

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 11:39:45 ; Search time 120.24 Seconds  
(without alignments)  
1738.900 Million cell updates/sec

Title: US-09-917-376-3

Perfect score: 4036

Sequence: 1 ATTQPTWSNVAIGGGFVD.....YIGTNGRGIYVGDIGGAPSG 740

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_29Jan04.\*

1: Geneseq1980s.\*

2: Geneseq1990s.\*

3: Geneseq2000s.\*

4: Geneseq2001s.\*

5: Geneseq2002s.\*

6: Geneseq2003as.\*

7: Geneseq2003bs.\*

8: Geneseq2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4036	100.0	740	6	ABP73016 Amino aci
2	4036	100.0	740	7	Add22926 Acidother
3	4036	100.0	740	7	Add22923 Acidother
4	4036	100.0	957	6	ABP73015 Amino aci
5	4036	100.0	957	6	Add22921 Acidother
6	2063	51.1	940	6	ABP99489 Amino aci
7	1680	41.6	726	7	Add22927 Aspergill
8	1579	39.1	838	7	Add42055 Trichoder
9	841.5	20.8	555	7	Add24919 Xanthomon
10	238	5.9	2468	6	ABU38411 Protein e
11	238	5.9	2468	6	ABP59933 Microbia
12	199.5	4.9	2435	6	ABU19529 Protein e
13	181.5	4.5	3892	6	ADA34216 Acinetoba
14	181	4.5	1308	6	ABU16642 Protein e
15	170	4.2	1074	6	ABU22692 Protein e
16	168.5	4.2	1439	6	ABU16643 Protein e
17	168.5	4.2	5291	7	ADC01014 Enterohae
18	167.5	4.2	1119	7	ABW01170 Candida a
19	166	4.1	1468	7	Add48744 Rat Prote
20	163.5	4.1	1289	6	ABU16977 Protein e
21	163	4.0	1684	2	AAR14948 Bacterial
22	160.5	4.0	3716	6	ABM15900 Mycobacte
23	159.5	4.0	1291	2	AAW59912 Amino aci
24	158.5	3.9	774	2	AAW35390 Flavobact
25	158.5	3.9	1440	5	ABB54801 Lactococc

ALIGNMENTS

RESULT 1

ABP73016

ID ABP73016 standard; peptide; 740 AA.

XX AC ABP73016;

XX XX

XX DT 03-JUN-2003 (first entry)

XX XX

XX DE Amino acid sequence of the avicelase AvIII catalytic domain.

XX XX

XX KW Avicelase; AvIII; glycoside hydrolase; enzyme; cellulase; biofuel;  
KW KW detergent; pulp processing; paper processing; feed processing; textile;  
KW KW cellulose.

XX OS Acidothermus cellulolyticus.

XX XX

XX PN WO2003012090-A2.

XX XX

XX PD 13-FEB-2003.

XX XX

XX PF 28-JUL-2001; 2001WO-US023818.

XX XX

XX PR 28-JUL-2001; 2001WO-US023818.

XX XX

XX PA (MIDE ) MIDWEST RES INST.

XX XX

XX PI Ding S, Adney WS, Vinzant TB, Himmel ME;

XX XX

XX DR WPI; 2003-248177/24.

XX XX

PT New thermostable AvIII peptide from Acidothermus cellulolyticus, useful  
PT for degradation of cellulose or in generating anti-AvIII antibodies for  
PT purifying recombinant AvIII polypeptides from genetically engineered  
PT host cells.

XX PS Claim 6; Page 8; 4pp; English.

XX XX

CC The present sequence is derived from a thermostable avicelase, designated  
CC AvIII. AvIII is a member of the glycoside hydrolase family of enzymes,  
CC and is a cellulase. AvIII is useful in the conversion of biomass to  
CC biofuels and biofuel additives. It may be useful in the production of  
CC detergents, pulp and paper processing, food and feed processing and in  
CC textile processes. The thermostable AvIII peptide is useful in the  
CC degradation of cellulose, and in generating specific anti-AvIII  
CC antibodies that are useful in purifying recombinant AvIII polypeptides  
CC from genetically engineered host cells, in detecting AvIII polypeptide  
CC expression, as well as a reagent tool for characterizing the molecular  
CC actions of the polypeptide. The AvIII polynucleotide is useful as a

CC source of probes or primers in various diagnostic assays

XX SQ Sequence 740 AA;

Query Match 100.0%; Score 4036; DB 6; Length 740;

Best Local Similarity 100.0%; Pred. No. 8.4e-271; Indels 0; Gaps 0;

Matches 740; Conservative 0; Mismatches 0; Mismatches 0; Mismatches 0;

Qy 1 ATTQPTWSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 60

Db 1 ATTQPTWSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 60

Qy 61 NNWYGVVSTAAADPINTNKWAAVGMVYNSDGDGAILRSSDQATWQITPLPFKLG 120

Db 61 NNWYGVVSTAAADPINTNKWAAVGMVYNSDGDGAILRSSDQATWQITPLPFKLG 120

Qy 121 NMPGRGGERLAVDPNDNIIYFCAPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDT 180

Db 121 NMPGRGGERLAVDPNDNIIYFCAPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDT 180

Qy 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNNPVFWSRDGATWQAVGAPTGF 240

Db 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNNPVFWSRDGATWQAVGAPTGF 240

Qy 241 IPHKGVPDPVNHVLIYATNSGTGGPYDGSGBVWKFVSTGVTWTRISPVSTDTANDYFGY 300

Db 241 IPHKGVPDPVNHVLIYATNSGTGGPYDGSGBVWKFVSTGVTWTRISPVSTDTANDYFGY 300

Qy 301 SGLTIDRQHPNTIMVATQISWNPDTIIFRSTDDGATWRIWDWTSYPNRSRYVLDISAE 360

Db 301 SGLTIDRQHPNTIMVATQISWNPDTIIFRSTDDGATWRIWDWTSYPNRSRYVLDISAE 360

Qy 361 PWLTFGVQPNPVPSPKLGWDEAMADPNSDRMLYGTGATLYATNDLTWDSGGQIHI 420

Db 361 PWLTFGVQPNPVPSPKLGWDEAMADPNSDRMLYGTGATLYATNDLTWDSGGQIHI 420

Qy 421 APVKGLEETAANDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSPVFTTGSVDY 480

Db 421 APVKGLEETAANDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSPVFTTGSVDY 480

Qy 481 AELNPSIIVRAGSDPSPQNDRIHVAFTDCKWQFQSEGGVTTGTTVAASADGSRFV 540

Db 481 AELNPSIIVRAGSDPSPQNDRIHVAFTDCKWQFQSEGGVTTGTTVAASADGSRFV 540

Qy 541 WAPGDPQPVVYAVFGNSWAASQGVANAGIISDRVNPKTFFYALNCTFFYRSTDDGVT 600

Db 541 WAPGDPQPVVYAVFGNSWAASQGVANAGIISDRVNPKTFFYALNCTFFYRSTDDGVT 600

Qy 601 QPVAAGLPSSGAVGMFHAVPGKEGDLWLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 660

Db 601 QPVAAGLPSSGAVGMFHAVPGKEGDLWLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 660

Qy 661 KSAPGSSYPVAVVGTGGTGAIRSDDCGTTWLVINDDQHOYGNWQAITGDHANLRV 720

Db 661 KSAPGSSYPVAVVGTGGTGAIRSDDCGTTWLVINDDQHOYGNWQAITGDHANLRV 720

Qy 721 YIGTNGRGIVYDGGAPSG 740

Db 721 YIGTNGRGIVYDGGAPSG 740

RESULT 2

ADD22926

ID ADD22926 standard; protein; 740 AA.

XX AC

XX ADD22926;

XX DT

XX 15-JAN-2004 (first entry)

XX DE

XX Acidothermus cellulolyticus avicelase AvIII catalytic domain.

XX KW

XX enzyme; AvIII; cellulose reduction; agricultural biomass;

XX municipal solid waste; glycoside hydrolase; avicelase.

XX

XX

XX

XX

XX Acidothermus cellulolyticus.

XX US2003108988-A1.

XX PD 12-JUN-2003.

XX PF 18-OCT-2002; 2002US-00155400.

XX PR 28-JUL-2001; 2001US-00917376.

XX PA (DING/) DING S.

XX PA (ADNEY/) ADNEY W S.

XX PA (VINZ/) VINZANT T B.

XX PA (HIMM/) HIMMEL M E.

XX PI Ding S, Adney WS, Vinzant TB, Himmel ME;

XX WPI; 2003-810853/76.

XX New isolated thermal tolerant avicelase polynucleotide useful for detection of a polynucleotide encoding AvIII and for reducing cellulose in a starting material, e.g. municipal solid waste.

XX Example 2; SEQ ID NO 6; 29pp; English.

XX The invention relates to an isolated polynucleotide molecule encoding a thermostable AvIII polypeptide. The polynucleotide is useful for detection of a polynucleotide encoding AvIII. The polynucleotide is useful for reducing cellulose in a starting material which involves administering to the starting material, e.g. agricultural biomass or municipal solid waste, a polypeptide molecule of the polynucleotide. The method further comprises administering a second polypeptide molecule chosen from the glycoside hydrolase family of proteins. The present sequence represents the amino acid sequence of Acidothermus cellulolyticus avicelase AvIII catalytic domain.

XX SQ Sequence 740 AA;

Query Match 100.0%; Score 4036; DB 7; Length 740;

Best Local Similarity 100.0%; Pred. No. 8.4e-271; Indels 0; Gaps 0;

Matches 740; Conservative 0; Mismatches 0; Mismatches 0; Mismatches 0;

Qy 1 ATTQPTWSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 60

Db 1 ATTQPTWSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 60

Qy 61 NNWYGVVSTAAADPINTNKWAAVGMVYNSDGDGAILRSSDQATWQITPLPFKLG 120

Db 61 NNWYGVVSTAAADPINTNKWAAVGMVYNSDGDGAILRSSDQATWQITPLPFKLG 120

Qy 121 NMPGRGGERLAVDPNDNIIYFCAPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDT 180

Db 121 NMPGRGGERLAVDPNDNIIYFCAPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDT 180

Qy 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNNPVFWSRDGATWQAVGAPTGF 240

Db 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNNPVFWSRDGATWQAVGAPTGF 240

Qy 241 IPHKGVPDPVNHVLIYATNSGTGGPYDGSGBVWKFVSTGVTWTRISPVSTDTANDYFGY 300

Db 241 IPHKGVPDPVNHVLIYATNSGTGGPYDGSGBVWKFVSTGVTWTRISPVSTDTANDYFGY 300

Qy 301 SGLTIDRQHPNTIMVATQISWNPDTIIFRSTDDGATWRIWDWTSYPNRSRYVLDISAE 360

Db 301 SGLTIDRQHPNTIMVATQISWNPDTIIFRSTDDGATWRIWDWTSYPNRSRYVLDISAE 360

Qy 361 PWLTFGVQPNPVPSPKLGWDEAMADPNSDRMLYGTGATLYATNDLTWDSGGQIHI 420

Db 361 PWLTFGVQPNPVPSPKLGWDEAMADPNSDRMLYGTGATLYATNDLTWDSGGQIHI 420

Qy 421 APVKGLEETAANDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSPVFTTGSVDY 480

Db 421 APVKGLEETAANDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSPVFTTGSVDY 480

Db 421 APMVKGLEETAVNDLISPPSGAPLISALDGLGGTHADVTAVPSTIFTSPVFTTGTSDVY 480  
 Qy 481 AELNPSIIIVRAGSFDPSQPNDRHVAFSTDGKKNWFOGSEPPGGVTTGGTVAASADGSRFV 540  
 Db 481 AELNPSIIIVRAGSFDPSQPNDRHVAFSTDGKKNWFOGSEPPGGVTTGGTVAASADGSRFV 540  
 Qy 541 WAPGDPQPVVYAVFGNSWAASQGVPPANAQIRSDRVNPKTFVALSNGTFYRSTDGCVTF 600  
 Db 541 WAPGDPQPVVYAVFGNSWAASQGVPPANAQIRSDRVNPKTFVALSNGTFYRSTDGCVTF 600  
 Qy 601 QPVAAGLPSSGAVGVMFHAFVPGKEGDLWLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG 660  
 Db 601 QPVAAGLPSSGAVGVMFHAFVPGKEGDLWLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG 660  
 Qy 661 KSAFGSSYPVAVFVGTGGTGVYRSDDCGTTWVLINDDQHOYGNWGOAITGDHANLRRV 720  
 Db 661 KSAFGSSYPVAVFVGTGGTGVYRSDDCGTTWVLINDDQHOYGNWGOAITGDHANLRRV 720  
 Qy 721 YIGTNGRGIVYDGGAPSG 740  
 Db 721 YIGTNGRGIVYDGGAPSG 740

RESULT 3  
 ADD22923  
 ID ADD22923 standard; protein; 740 AA.  
 AC ADD22923;  
 XX  
 DT 15-JAN-2004 (first entry)  
 XX  
 DE Acidothermus cellulolyticus avicelase Aviii catalytic domain.  
 XX  
 KW enzyme; Aviii; cellulose reduction; agricultural biomass;  
 KW municipal solid waste; glycoside hydrolase; avicelase.  
 XX  
 OS Acidothermus cellulolyticus.  
 XX  
 PN US2003108988-A1.  
 XX  
 PD 12-JUN-2003.  
 XX  
 PF 18-OCT-2002; 2002US-00155400.  
 XX  
 PR 28-JUL-2001; 2001US-00917376.  
 XX  
 PA (DING/) DING S.  
 PA (ADNEY/) ADNEY W S.  
 PA (VINZ/) VINZANT T B.  
 PA (HIMM/) HIMMEL M E.  
 XX  
 PI Ding S, Adney WS, Vinzant TB, Himmel ME;  
 XX  
 DR WPI; 2003-810853/76.  
 XX  
 PT New isolated thermal tolerant avicelase polynucleotide useful for  
 PT detection of a polynucleotide encoding Aviii and for reducing cellulose  
 PT in a starting material, e.g. municipal solid waste.  
 XX  
 PS Claim 16; SEQ ID NO 3; 29pp; English.  
 XX  
 CC The invention relates to an isolated polynucleotide molecule encoding a  
 CC thermostable Aviii polypeptide. The polynucleotide is useful for  
 CC detection of a polynucleotide encoding Aviii. The polynucleotide is  
 CC useful for reducing cellulose in a starting material which involves  
 CC administering to the starting material, e.g. agricultural biomass or  
 CC municipal solid waste, a polypeptide molecule of the polynucleotide. The  
 CC method further comprises administering a second polypeptide molecule  
 CC chosen from the glycoside hydrolase family of proteins. The present  
 CC sequence represents the amino acid sequence of Acidothermus  
 CC cellulolyticus avicelase Aviii catalytic domain.  
 XX  
 SQ Sequence 740 AA;

Query Match 100.0%; Score 4036; DB 7; Length 740;  
 Best Local Similarity 100.0%; Pred. No. 8.4e-271;  
 Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTOPYTWNSVAIGGGFVDGIVFNEGAPGLIYVTRDTCGMWRDAANGRWITPLLDVWGV 60  
 Db 1 ATTOPYTWNSVAIGGGFVDGIVFNEGAPGLIYVTRDTCGMWRDAANGRWITPLLDVWGV 60  
 Qy 61 NNWGVNGVVSIAADPINTNKVAAVGMVTSNDPNDGAILRSSDQATWQITPLPPLKGG 120  
 Db 61 NNWGVNGVVSIAADPINTNKVAAVGMVTSNDPNDGAILRSSDQATWQITPLPPLKGG 120  
 Qy 121 NMPGRGEMERLAVDPNNNDNIIYFGAPSGKGLWRSTDGATWQMTNFPDVGTYIANPTDT 180  
 Db 121 NMPGRGEMERLAVDPNNNDNIIYFGAPSGKGLWRSTDGATWQMTNFPDVGTYIANPTDT 180  
 Qy 181 TGYQSDIQGVVWVAFDSSSSSLGQASKTIFGVADPNPNPVFWSRDGGATWQAVPGAPTGF 240  
 Db 181 TGYQSDIQGVVWVAFDSSSSSLGQASKTIFGVADPNPNPVFWSRDGGATWQAVPGAPTGF 240  
 Qy 241 IPHKGVEDPVNHLVLIATSNLTGGPYDGGSDGVWKFVSTGTRISPVSTDTANDYFGY 300  
 Db 241 IPHKGVEDPVNHLVLIATSNLTGGPYDGGSDGVWKFVSTGTRISPVSTDTANDYFGY 300  
 Qy 301 SGLTIDROHPNTIMVATOISWMPDPTIIIFRSTDDGATWTRINDWTSYPNRSRLRYLDISAE 360  
 Db 301 SGLTIDROHPNTIMVATOISWMPDPTIIIFRSTDDGATWTRINDWTSYPNRSRLRYLDISAE 360  
 Qy 361 PWLTFGVQPNPPVPSPLKGMWDEAMAIIDPFNSDRMLYGTGATLYATNDLTWKDSSGQIHI 420  
 Db 361 PWLTFGVQPNPPVPSPLKGMWDEAMAIIDPFNSDRMLYGTGATLYATNDLTWKDSSGQIHI 420  
 Qy 421 APMVKGLEETAVNDLISPPSGAPLISALDGLGGTHADVTAVPSTIFTSPVFTTGTSDVY 480  
 Db 421 APMVKGLEETAVNDLISPPSGAPLISALDGLGGTHADVTAVPSTIFTSPVFTTGTSDVY 480  
 Qy 481 AELNPSIIIVRAGSFDPSQPNDRHVAFSTDGKKNWFOGSEPPGGVTTGGTVAASADGSRFV 540  
 Db 481 AELNPSIIIVRAGSFDPSQPNDRHVAFSTDGKKNWFOGSEPPGGVTTGGTVAASADGSRFV 540  
 Qy 541 WAPGDPQPVVYAVFGNSWAASQGVPPANAQIRSDRVNPKTFVALSNGTFYRSTDGCVTF 600  
 Db 541 WAPGDPQPVVYAVFGNSWAASQGVPPANAQIRSDRVNPKTFVALSNGTFYRSTDGCVTF 600  
 Qy 601 QPVAAGLPSSGAVGVMFHAFVPGKEGDLWLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG 660  
 Db 601 QPVAAGLPSSGAVGVMFHAFVPGKEGDLWLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG 660  
 Qy 661 KSAFGSSYPVAVFVGTGGTGVYRSDDCGTTWVLINDDQHOYGNWGOAITGDHANLRRV 720  
 Db 661 KSAFGSSYPVAVFVGTGGTGVYRSDDCGTTWVLINDDQHOYGNWGOAITGDHANLRRV 720  
 Qy 721 YIGTNGRGIVYDGGAPSG 740  
 Db 721 YIGTNGRGIVYDGGAPSG 740

RESULT 4  
 ABP73015  
 ID ABP73015 standard; protein; 957 AA.  
 XX  
 AC ABP73015;  
 XX  
 DT 03-JUN-2003 (first entry)  
 XX  
 DE Amino acid sequence of the avicelase Aviii.  
 XX  
 KW Avicelase; Aviii; glycoside hydrolase; enzyme; cellulase; biofuel;  
 KW detergent; pulp processing; paper processing; feed processing; textile;  
 KW cellulose.  
 XX  
 OS Acidothermus cellulolyticus.





useful for reducing cellulose in a starting material which involves administering to the starting material, e.g. agricultural biomass or municipal solid waste, a polypeptide molecule of the polynucleotide. The method further comprises administering a second polypeptide molecule chosen from the glycoside hydrolase family of proteins. The present sequence represents the amino acid sequence of Acidothermus cellulolyticus avicelase Aviii.

Sequence 957 AA;

Query Match 100.0%; Score 4036; DB 7; Length 957;  
Best Local Similarity 100.0%; Pred. No. 1.2e-270;  
Matches 740. Conservative 0. Mismatches 0. Indels 0.

Qy	1	ATTOPYTWSNVAICGGGFVDGIVFNEGAPGLIYVRTDIGMYRWDAANGRWIPLLDWVGW	60
Db	47	ATTOPYTWSNVAICGGGFVDGIVFNEGAPGLIYVRTDIGMYRWDAANGRWIPLLDWVGW	106
Qy	61	NNWYNGVGVSIADDPINTNKWAAVGMWYTSWDPNDGAILRSSDQGA TWQITPLPFKLG	120
Db	107	NNWYNGVGVSIADDPINTNKWAAVGMWYTSWDPNDGAILRSSDQGA TWQITPLPFKLG	166
Qy	121	NMPGRGNGERLAVDPNNNDNIIYFCAPSGKGLWRSTDGATWSQMTNFPDVGTIYIANPTDT	180
Db	167	NMPGRGNGERLAVDPNNNDNIIYFCAPSGKGLWRSTDGATWSQMTNFPDVGTIYIANPTDT	226
Qy	181	TGYQSDIQGVVWVAFDXSSSLGQASKTI FVGVA DPNNPVFWSRDPGGATWQAVPGAPTGF	240
Db	227	TGYQSDIQGVVWVAFDXSSSLGQASKTI FVGVA DPNNPVFWSRDPGGATWQAVPGAPTGF	286
Qy	241	IPHKGVDPNNHVLIIATSNITGGBYDGS SCDGWKFSVTSGTWTIRISPVPSSTDANDYRGY	300
Db	287	IPHKGVDPNNHVLIIATSNITGGBYDGS SCDGWKFSVTSGTWTIRISPVPSSTDANDYRGY	346
Qy	301	SGLTIDRQHPNTIMVATQISWWPDTIIFRSTDGGATWTRIWDWTSYPNRSRYVLLDISAE	360
Db	347	SGLTIDRQHPNTIMVATQISWWPDTIIFRSTDGGATWTRIWDWTSYPNRSRYVLLDISAE	406
Qy	361	PWLTFGVQPNPPVPSPKLGWMD EAMAIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHI	420
Db	407	PWLTFGVQPNPPVPSPKLGWMD EAMAIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHI	466
Qy	421	APWKYGLI EETAVNDLIISPPSGAPLISALGD LGGTHADVTAVPSITFTSPVFTTCTSVDY	480
Db	467	APWKYGLI EETAVNDLIISPPSGAPLISALGD LGGTHADVTAVPSITFTSPVFTTCTSVDY	526
Qy	481	AELNPSIIVRAGSPDPSPQPNDRHVA FSTDCGKNWFGQSEPGCVTTGGTVAASADGSRFV	540
Db	527	AELNPSIIVRAGSPDPSPQPNDRHVA FSTDCGKNWFGQSEPGCVTTGGTVAASADGSRFV	586
Qy	541	WAPGDPQPVVYAVFGNSWMAASQGV PANAQIRSDRVNPKTFYALNSGTHFYKSTDGGVTTF	600
Db	587	WAPGDPQPVVYAVFGNSWMAASQGV PANAQIRSDRVNPKTFYALNSGTHFYKSTDGGVTTF	646
Qy	601	QPVAAGLPSSGAGVGMFHVPGKSGDLWLAASSGLYHSTNGC SSWSAITGVSSS VNVVFG	660
Db	647	QPVAAGLPSSGAGVGMFHVPGKSGDLWLAASSGLYHSTNGC SSWSAITGVSSS VNVVFG	706
Qy	661	KSAPGSSYPAVFVVGTTIGGVTGAYRSDDCGTTWVLLINDDQHQYGNWGQAITGDHANLRV	720
Db	707	KSAPGSSYPAVFVVGTTIGGVTGAYRSDDCGTTWVLLINDDQHQYGNWGQAITGDHANLRV	766
Qy	721	YIGTNGRGIYVGDITGGAPSG 740	
Db	767	YIGTNGRGIYVGDITGGAPSG 786	

## RESULT 6

ABB99489

ID ABB99489 standard: protein: 940 AA:

XX

AC ABB99489;

DT	03-MAR-2003	(first entry)	
XX	Amino acid sequence of a xyloglucanase enzyme.		
XX	Xyloglucanase; enzyme; family 74; glycosyl hydrolase; cellulosic fiber;		
KW	textile scouring; cellulose fiber processing; ratting.		
KW	Jonesia sp.		
OS	W0200277242-A2.		
XX	W0200277242-A2.		
XX	03-OCT-2002.		
XX	27-MAR-2002; 2002WO-DK000210.		
PF	27-MAR-2001; 2001DK-00000504.		
XX	(NOVO ) NOVOZYMES AS.		
PA	Duffner F, Sjöholm C;		
XX	WPI; 2003-092855/08.		
XX	N-PSDB; ABV76941.		
DR			
DR			
XX	New xyloglucanase enzyme belonging to family 74 of glycosyl hydrolases,		
PT	and endogenous to a bacterium, useful in the textile industry for		
PT	improving properties of cellulosic fibers, yarn or fabric.		
XX	Claim 5; Page 72-75; 76pp; English.		
PS			
XX			
CC	The present sequence represents a xyloglucanase enzyme, belonging to		
CC	family 74 of glycosyl hydrolases. The enzyme is isolated from Jonesia sp.		
CC	DSM14140. The enzyme is useful in processes for machine treatment of		
CC	fabrics. It is also useful in the textile industry for improving the		
CC	properties of cellulosic fibers, yarn, woven or non-woven fabric, and in		
CC	a textile scouring process step. The xyloglucanase enzyme is also useful		
CC	in the cellulose fiber processing industry for ratting of fibers such as		
CC	hemp, jute, flax and linen. It is useful for preventing binding of		
CC	certain soils to the xyloglucan left on the cellulosic material		
XX			
SQ	Sequence 940 AA;		
	Query Match	51.1%;	Score 2063; DB 6; Length 940;
	Best Local Similarity	52.1%;	Pred. No. 5.2e-134;
	Matches	386;	Conservative 112; Mismatches 225; Indels 18; Gaps 9;
Qy	6	YTSWSNVAIGGGFVDGIVFNEGAPGLYVRTDGGMYRWDAANGRWIPLLDWGVNNWGY	65
Db	44	YSNSNBEIVGGFVPGIVFNGKDPGLVIARTDGGAYRLNDSTGRWIPPLTDHIGWDDWSH	103
Qy	66	NGVYSIAADPINTNKVAAVGMVTSNDPNDGAILRSSDQAGATWITPLPFKLGNNMGR	125
Db	104	SGILSLATDPVDINRVYLAAGTYSMDWDPQNGAILRSADKGETWEKTMLEPRVGGNNMGR	163
Qy	126	GMGERLAVDPNNNNILYFGAPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDTTG-YQ	184
Db	164	GMGERLAVDPNNNNKLYLFGAESNGLWKSTDYDKTGWKVTSPFNAGNYVA--DASGAYT	220
Qy	185	SDIQGVVWVAFDKSSSLGQASKTIIFVGVAADPNPNVFWSDGGGATQVAGPATGPTPHK	244
Db	221	QNGQGVVWVTFDPTSAKAGTTOTIYVGVADAKQNNVYRSTDGGATWQRVFGQPTGFLAQK	280
Qy	245	GVDPNNHVLVIATSNVTGGPYDGGSDGVWKFVTSVTGWTWTRISVPVSTDTANDVFGYSGLT	304
Db	281	GVLDHKKQQQLYIATSDTGGPYDGGSDGVWRLDLSGGQWTRISPIPST-SNSAFGYSGLA	339
Qy	305	IDROHPNTIMVATQISWNPDTIIFRSTDGATWTRIMDWTSYPNRSRLRYVLDISAEPLT	364
Db	340	IDRKNPDTIMVQSVSWPDMYVYRSTDCKTWSPIWELNGSQPRTKQYNDHYSGAPWLD	399
Qy	365	FGVQNPFPVPSPLGWMDAEMADIPFNSDRMLYGTGATLYATNDLTWKWDSGGQIHIAPMV	424
Db	400	FGNTAKEPEANPLKGMWTSQFEDPHNSDRFFYGTGAGIYCGTNLTINWDKGGKYVDITVKA	459



ethanol production; detergent composition; fabric treatment;  
textile treatment; enzyme.

Hypocrea jecorina.  
WO2003089598-A2.  
30-OCT-2003.  
17-APR-2003; 2003WO-US011831.  
19-APR-2002; 2002US-0373987P.  
(NOVO) NOVOZYMES BIOTECH INC.  
Michael R. Zaretsky E, Haas J;  
WPI; 2003-845528/78.  
N-PSDB; ADD42054, ADD42060.  
New polypeptides having Family 74 xyloglucanase activity, and encoding  
nucleic acid molecules, useful for degrading cellulose- and hemicellulose  
-containing biomass to ethanol or as a detergent.  
Claim 8; SEQ ID NO 2; 96pp; English.  
The invention relates to a Family 74 xyloglucanase (ADD42055) from the  
fungus *Trichoderma reesei* (Hypocrea jecorina), and nucleic acids encoding  
it (e.g., ADD42054). The enzyme hydrolyses the beta-1,4-glycosidic  
linkages in the backbone of xyloglucan to xyloglucan oligosaccharides.  
The invention also relates polypeptide sequences at least 70% identical  
to the enzyme, expression vectors and host cells comprising a nucleic  
acid of the invention, the recombinant production of the enzyme, and  
mutant enzymes and the nucleic acids encoding them. The xyloglucanase of  
the invention can be used in the degradation of cellulose- and  
hemicellulose-containing biomass to produce ethanol. It can also be used  
in a detergent composition for treating fabric during a machine washing  
cycle. The present sequence represents the *Trichoderma reesei* Family 74  
xyloglucanase.

Query Match 39.1%; Score 1579; DB 7; Length 838;  
Best Local Similarity 43.0%; Pred. No. 1.5e-100;  
Matches 325; Conservative 123; Mismatches 264; Indels 44; Gaps 20;

6 YTSNVAI-GGGFVDFGIVNEGAPGILYVFTDGGMYRWDAAANGRIPLDWW-...GW 60  
21 FSKNVKLGCGGFGVPGIIFHPKTKGVAYARTDIGLYRLN-ADDSWATVTDGIADNAGW 79  
61 NNWYNGVVSIAADPINTNKVAAVGMVYNSWDNDGAILRSSDQATWQITPLPKLG 120  
80 HNW---GIDVALDPQDDQKYAAVGMVYNSWDNDGAILRSSDQATWQITPLPKLG 136  
121 NMPGRGGERLAVDPNNNTLYFCAPSGKGLWRGTSQATWSOMTNPDPVGTYIANPTDT 180  
137 NMPGRGGERLAVDPANNIYIFGARGNGLWKSTDGVTFSKVSSTATGTIYIPDFS 196  
181 TGYQSDIQGVVWVAFDKSSSLGQASKTI FVGVA DP-NNPVFWSRGGATWQAVPGAPT 239  
197 NGYNSDKQGLWVTFDSTSTGATGSRIFVGTADNITASVYVSTNAGSTWSAVFGPGK 256  
240 FIPKGVDFPNVHLYIATNTGPGYDGSQDVWKFVSTGWTWRIIPVSTDTANDYFG 299  
257 YFPHKAKLQPAEKALYLYTSDGTGPGYDGLGWSVRYDIAGGTWKTIDITVSGSDL-...YFG 313  
300 YSLGTTDROHPNTIMVATQISWPDITIFRSTDCGATWTRIDWTSYPNRSLRVLDISA 359  
314 FGLGLDLQKPLVAVLASLNSWPDQAQFRSTDSGTTWSPINAWASTPTETYYISITPK 373  
360 EPMWTFG-VQNPFPVPS-...PKLGMDEAMAIDPFNSDRMLYGTGATLYATNDLTKWDS 414  
374 APWKNNFIDVTSESPDGLIKRLGWMIESLEIDPTDSNHWLYGTGTFIAGHDLTNDWT 433

QY 415 GQIHIAPMVKLEETAVNDLISPPSGAPLISALGLDGGFTHA---DVTAVPSTIFTSPV 471  
DB 434 RHNVSIQSLADGIEEFSVQDLASAPGSELLAAVGDNDGFTFASRNDLGTSPQVWATPT 493  
QY 472 FTTGTSVDYAEINPSIIVRAGSDFDSSQPNDRHVAFTDGGKWFQGSSEPGVTTGGTVA 531  
DB 494 WATSTSDYAGNSVSVVGVNVTAGTQQ-----VAISSDGGATWSIDYAADTSMNGGTVA 548  
QY 532 ASADGSRFWAPDGPQPVVYAVGVGNSWAASQGVPAQAIRSDRVNPKTFYALSNGTFF 591  
DB 549 YSADGDTILWSTASSG---VQRSQFGSPASVSSLPAGAVIAADKKTNSVFYAGSGSTFF 605  
QY 592 RSTDCGVTFPVAAAGLPSSGAVGVN---FHAVPGKEGDLMLAASSGLYHSTNGGSSWAI- 648  
DB 606 VSKDTGSSP---TRG-PKLGSA GTIRDIAAHPPTAGTLYVSTDVGFIRSTDSGTTFGQVS 661  
QY 649 TGVSSAVNVGFKSAPGSSYPVAVFVGTIGGVTGA---YRSDDCGTTWLVLLINDHQHGVNW 706  
DB 662 TALNTYQIALGVGS-GSNW-NLYAFGT--GPSGARLYASGDSGASWTDIQSGQFGSID 717  
QY 707 GQAITGDHANLRVVGITNGRGIIV--GDIGGAPSG 740  
DB 718 STKVAGSGTAGQVYVGVNNGRGVFAQGTGVTGGTGG 753

RESULT 9  
ADD24919  
ID ADD24919 standard; protein; 555 AA.  
XX  
AC ADD24919;  
XX  
DT 15-JAN-2004 (first entry)  
XX  
DE Xanthomonas campestris cellulase #1.  
XX  
KW Directed genetic engineering; galactomannanase; reduced activity;  
KW enhanced activity; xanthan gum production; suspension stability;  
KW emulsion stability; temperature resistance; pseudoplasticity; amylase;  
KW cellulase; extracellular protease; intracellular protease;  
KW glucose dehydrogenase; enzyme.  
XX  
OS Xanthomonas campestris.  
XX  
FN US2003036176-A1.  
XX  
PD 20-FEB-2003.  
XX  
PF 10-AUG-2001; 2001US-00927827.  
XX  
PR 28-MAR-2001; 2001US-0279493P.  
XX  
PA (BOWE/) BOWER S G.  
PA (RAMS/) RAMSEIER T M.  
XX  
PI Bower SG, Ramseier TM;  
XX  
XX WPI; 2003-625389/59.  
DR N-PSDB; ADD24893.  
XX  
PT New transformed cell or organism having reduced or enhanced activity of  
PT at least one protein, useful for producing xanthan gum, which are useful  
PT for providing formulations and properties.  
XX  
PS Claim 1; SEQ ID NO 47; 135pp; English.  
XX  
CC The present invention relates to polypeptide and polynucleotide sequences  
CC from *Xanthomonas campestris* which may be used for activity reduction or  
CC enhancement using directed genetic engineering. A transformed cell or  
CC organism having reduced or enhanced activity of at least one such protein  
CC e.g. galactomannanase can be generated by disrupting the gene encoding  
CC the protein. The activity of the protein is reduced by the presence of an  
CC antisense nucleic acid sequence. The nucleic acid sequence of the gene

CC encoding the protein is a recombinant sequence having at least one  
 CC mutation as compared to the wild-type gene encoding the protein. The  
 CC transgenic cell or microorganism are useful for producing xanthan gum,  
 CC which are useful for providing formulations and properties, such as long-  
 CC term suspension and emulsion stability in alkaline, acid, and salt  
 CC solutions, temperature resistance, and pseudoplasticity. The present  
 CC sequence represents an enzyme relating to the present invention.  
 XX  
 SQ Sequence 555 AA;

Query Match 20.8%; Score 841.5; DB 7; Length 555;  
 Best Local Similarity 37.1%; Pred. No. 1e-49;  
 Matches 186; Conservative 81; Mismatches 193; Indels 41; Gaps 12;  
 QY 1 ATTPQYTSNNVAIGGGGVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLDITWGW 60  
 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 60  
 73 ATSGPYQWRSVAIGGGGFTVGLFHPAERGLAYARTDVGAYRWDAQAQQTALTDWGA 132  
 QY 61 NNWYNGVVSAAADPINTNKVAAVGMTNSWDNDGAILRSSDGGATWQTITPLPFKLG 120  
 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 120  
 133 DQNLMLGIDAFVDPADADALYLAAGTYMHE-RAGTAALVLRSENRGRTFERADLPFKLG 191  
 QY 121 NMPGRGGERLAVDPNNNLYFGAPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDT 180  
 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 180  
 192 NQLGRANGERLAVDPHGRVILLGRDA-GLWRSDRGAAHWAKVASFPDAAALAGATARNH 250  
 QY 181 TGYOSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNNPFWRSRGGATWQAVGAPTGF 240  
 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 240  
 251 VGREQAV-GIAFVVFDAAGNGNSPTPRIYGVSTEQTSLYVSEDAGSWAPVAGQPRGL 309  
 QY 241 IPHKGVPDPVNVHLYATNSNTGGPYDGSDDWKPSTGVTWTRISPVSTDTTANDYFGY 300  
 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 300  
 310 RPSHMAGSGDH-WYLSYCDQPGDLMAGGALWKFTPAQGRWRREISPIQP-PASGDGFGW 367  
 QY 301 SGLTIDRQHPNTIIVATQISWWPDTIIPRSYDGGATWTRINDWTSYNSRSLRYVLDIAE 360  
 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 360  
 368 GAVAVDPHQPOLLASTFRRTPRDELRYSDVDGKHHTPL-----LADAVFDHSA 418  
 QY 361 PWLTFGVQPNPPVSPKLGWMDAMADIPFNSDRMLYGTGATLVATNDLTKWDSGGQHI 420  
 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 420  
 419 PW-TAHATPH-----NWG-ALADIPFGNHALFTVGYIWARNL-----QDFA 460  
 QY 421 AP-----MVKLEETAVNDLISPPSGAPLISALGDLGGFTHADVAVPSTIFTSPVF 472  
 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 472  
 461 APQRPQWQFQDGLLEETVPLDLLSPMAGALLSALGDLIDGFRHDDLDRV-QLQVAGPRL 519  
 QY 473 TTGTSVDYAEINPSIIVRAGS 493  
 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 493  
 520 TNGESIDAAGQAPQWVRSGT 540

RESULT 10  
 ABU38411  
 ID ABU38411 standard; protein; 2468 AA.  
 XX  
 AC ABU38411;  
 XX  
 DT 19-JUN-2003 (first entry)  
 XX  
 DE Protein encoded by Prokaryotic essential gene #23938.  
 XX  
 KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 XX  
 OS Pseudomonas aeruginosa.  
 XX  
 FN WO200277193-A2.  
 XX  
 PD 03-OCT-2002.  
 XX  
 PF 21-MAR-2002; 2002WO-US009107.  
 XX  
 PR 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX  
 DR WPI: 2003-029926/02.  
 DR N-PSDB; AC422281.  
 XX  
 PT New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 PS  
 XS Claim 25; SEQ ID NO 63335; 1766pp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 2468 AA;  
 Query Match 5.9%; Score 238; DB 6; Length 2468;  
 Best Local Similarity 22.8%; Pred. No. 4.8e-07;  
 Matches 205; Conservative 88; Mismatches 314; Indels 292; Gaps 50;  
 QY 11 VAIGGGGVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLDITWGNWNGY----- 65  
 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 65  
 1084 VNLSSGSLG---TAEPGSTVILTD-----GNGNPFAEVTADSGSNWTFPSTP 1130  
 QY 66 --NG-VWSIADPINTNKVAAVGMTNS-----WDPNDGAILR-SSDGGATWQITPLP 115  
 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 115  
 1131 TANGTVNVVAQDASGNSPPATVTVDSAPAPVNPNSGVVISCTAEGATVLT--- 1187  
 QY 116 FKLGNMPPGRGGERLAVDPNNNLYFGAPSGKGLWRSTDGATWSQMTNFPDVGTYIA 175  
 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 175  
 1188 -DAGGNPIQGVTD-----GSGNWSFTP-----GTPLA 1214  
 QY 176 N-----PTDTTGYQS-----DIQGVVWVAFKSSSL--GQASKTIFGVADPN-N 218  
 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 218  
 1215 NGTVIVATATPTGNTGTPQAATVDAVAPPVDPFNSNGTTISGTAEAGAKVILTDGNGN 1274  
 QY 219 PV-FWSRDGGATWQAVPGAP-----TG-----FIPHKGVDPVN 251

Db 1275 PIGETTADSGNWSFTPGTPLANGTVVNAVAQDPAGNTGPGSTTVDAVAPNTPVVNPNS 1334  
 Qy 252 HVLVIATNTGGPY---DGSSGDVWKFVSVT-SGTW--TRISPVPS-----TDTANDYFGYS 301  
 Db 1335 GNLLNGTAEPGSTVTLTDGNGNPIGQTTADSGNWSFTPGSQLPNGTVVNVVITASDAAGNT 1394  
 Qy 302 GL-----TIDRQHPNTIMV-----ATQISWMPD---TIIFRSTDDG-----ATWTRIDW 342  
 Db 1395 SLPATTIVDSSLSIPQVDPNSGVSIGTADAGNTIIII--TDGNGNPIGQTTADSGNWS 1452  
 Qy 343 WT-----SYNRSRLRYVLDISAEPLW--TFGVQPNPPVPSPKLGWDEMAIDPFNS 392  
 Db 1453 FTPGIPLPDGTVVNVVARSPPNVDSPAVITVDGVAAPV-----IDPSNG 1499  
 Qy 393 DRMLYGT--GATLYATNDLTKWDSGGQIHIAPMVKGLEETAVNDLISP---PSGAPLI 445  
 Db 1500 TE-ISGTAEAGATVILT-----DGGN-----PIGQATADSGNWTFTPTPLANGTVIN 1548  
 Qy 446 SALGDLGGFTH-----ADVTAVPSTIF--TSPVFTTGTSDVVAELNPSIIVRAGSPDPS 497  
 Db 1549 AVAQDPAGNTSGPASVTVDAIAPPAPVINPNSGVVISGT----AEAGATVILTIDGNGNPI 1604  
 Qy 498 SQPNDRHFAFSTDGKNWF-----QSEPPGGVT----- 525  
 Db 1605 GQ-----VTADSGNWSFTPGTPLANGSVINALAQDAAGNNSPTSATVDSLAPAPV 1657  
 Qy 526 ----TGGTVAASADGSRFWAPGDPQPV--VYAVGFNSWAASQGVPANAIQSRDRVNP 579  
 Db 1658 IDPSNGSVIAGTAEGATVILTIDGNGNPIGQTTADSGN--WSFTPGTP----- 1704  
 Qy 580 KTFVALSNGTFYR-----STDGGV--TFQPVAAAGLP---SSGAVGVMPHAPVKGED 626  
 Db 1705 ----LSNGTVVNAVAQDAAGNTSGPVSTTVDAVAPATVIDPNSGVLSGTAEFGVRVI 1759  
 Qy 627 LMLAASGLYHS--TNGSSSIAITGV-----SSAVNVGFGKSAPGSSYPVAVVGTIGGVT 681  
 Db 1760 LTDGNGNPIGQTLADSGNWSFTPGTPLANGTVVNAVAQDPAGNTSGPASITVDTVAPAT 1819  
 Qy 682 GAYRSDDCGTTWVINDDQHOYGNWGAQITGDHANLRVYIGTNGRGIVYGDIGGAPSG 740  
 Db 1820 -----PVINPSN-----GSVITGT-AEVGAKVILTIDGNGNPIGETTADSGS 1859

## RESULT 11

ABP59933  
 ID ABP59933 standard; protein; 2468 AA.

AC ABP59933;

DT 28-AUG-2003 (first entry)

XX Microbial resistance gene PA1874 protein.

DE Biofilm; microbial resistance.

KW Unidentified.

OS WO2003041483-A2.

PN 22-MAY-2003.

PD 18-SEP-2002; 2002WO-US029565.

XX 18-SEP-2001; 2001US-0323241P.

PR (DART-) DARTMOUTH COLLEGE.

PA O'toole GA, Mah T;

PI WPI; 2003-468567/44.

XX N-PSDB; ACC59398.

XX

PT Identifying modulators of microbial resistance of organisms in biofilms,  
 PT e.g. inhibitor of biofilm formation, by employing expression controls, or  
 PT efflux pumps containing polypeptides, of genes associated with biofilm  
 PT resistance.  
 XX Claim 33; Fig 7; 102pp; English.

PS Claim 33; Fig 7; 102pp; English.

XX The present invention relates to a method of identifying a compound  
 CC capable of altering the sensitivity of a microorganism to an  
 CC antimicrobial agent by employing efflux pumps comprising polypeptides  
 CC encoded by the following genes: PA1874, PA4142, PA2389, PA1876, PA4143,  
 CC PA390 or PA1163. The method is useful for identifying modulators of  
 CC microbial resistance of an organism in a biofilm. The methods are also  
 CC useful for identifying genes that encode proteins that play a role in  
 CC biofilm resistance. The method is particularly useful for screening  
 CC compounds or discovering compositions that will inhibit biofilm formation  
 CC and overcome their resistance mechanisms. These methods are particularly  
 CC useful in medical, industrial or natural settings, where formation of  
 CC biofilms can have serious negative consequences and result in high costs  
 CC both in human health and economic terms. The present sequence is a  
 CC protein shown in the exemplification of the invention

XX Sequence 2468 AA;

Query Match 5.9%; Score 238; DB 6; Length 2468;  
 Best Local Similarity 22.8%; Pred. No. 4.8e-07;  
 Matches 205; Conservative 88; Mismatches 314; Indels 292; Gaps 50;

Qy 11 VAIGGGFVDFIVNEGAPGLIYVTDIGMYRDAAGRWIPLLDWVGNWGY----- 65  
 Db 1084 VNLSNGSSLSG---TAEFGSTVILT-----GNGNPIAEVTADSGNWTPTSTP 1130

Qy 66 --NG-VVSIADPINTNKVAAVGYMYS-----WDPNDGAILR--SSDQATWQITPLP 115  
 Db 1131 IANGTVVNVVNAQDAGNSPPATVTVDSAPPAPVINPNSGVVISGTAEAGATVLT--- 1187

Qy 116 FKLGGMGRGGERGLAVDPNNDNILYFGAPSGKGLMRSTDGATWSQMTNFPDVGTYIA 175  
 Db 1188 -DAGNPIGQVTAD-----GSGNWSFTP-----GTPLA 1214

Qy 176 N-----PTDTGYQS-----DIQVVVAVFKSSSL--GQASKTIFVGVADPN-N 218  
 Db 1215 NGTVIVATADPTGNTGPOAATTVDAVAPPVDPISNGTTISGTAEAGAKVILTIDGNGN 1274

Qy 219 PV-FWSRDGGATWQAVPGAP-----TG-----FIPHKGVDPPVN 251  
 Db 1275 PIGETTADSGNWSFTPGTPLANGTVVNAVAQDPAGNTGPGSTTVDAVAPNTPVVNPNS 1334

Qy 252 HVLVIATNTGGPY---DGSSGDVWKFVSVT-SGTW--TRISPVPS-----TDTANDYFGYS 301  
 Db 1335 GNLLNGTAEPGSTVTLTDGNGNPIGQTTADSGNWSFTPGSQLPNGTVVNVVITASDAAGNT 1394

Qy 302 GL-----TIDRQHPNTIMV-----ATQISWMPD---TIIFRSTDDG-----ATWTRIDW 342  
 Db 1395 SLPATTIVDSSLSIPQVDPNSGVSIGTADAGNTIIII--TDGNGNPIGQTTADSGNWS 1452

Qy 343 WT-----SYNRSRLRYVLDISAEPLW--TFGVQPNPPVPSPKLGWDEMAIDPFNS 392  
 Db 1453 FTPGIPLPDGTVVNVVARSPPNVDSPAVITVDGVAAPV-----IDPSNG 1499

Qy 393 DRMLYGT--GATLYATNDLTKWDSGGQIHIAPMVKGLEETAVNDLISP---PSGAPLI 445  
 Db 1500 TE-ISGTAEAGATVILT-----DGGN-----PIGQATADSGNWTFTPTPLANGTVIN 1548

Qy 446 SALGDLGGFTH-----ADVTAVPSTIF--TSPVFTTGTSDVVAELNPSIIVRAGSPDPS 497  
 Db 1549 AVAQDPAGNTSGPASVTVDAIAPPAPVINPNSGVVISGT----AEAGATVILTIDGNGNPI 1604

Qy 498 SQPNDRHFAFSTDGKNWF-----QSEPPGGVT----- 525  
 Db 1605 GQ-----VTADSGNWSFTPGTPLANGSVINALAQDAAGNNSPTSATVDSLAPAPV 1657

Qy 526 ----TGGTVAASADGSRFWAPGDPQPV--VYAVGFNSWAASQGVPANAIQSRDRVNP 579

Db 1658 IDPNSGVIACTAGATVILTDGNGNPVGQVTDGSGN-WSFTGTP----- 1704  
Qy 580 KTFYALNGTFYR-----STDGGV--TFQPVAAGLP---SSGAVGVMFHAPVKGED 626  
Db 1705 -----LSNGTVNAVAQDAAGTSGPVSTTVDAVAPATPVIDPSNGVLSGTAEFGVRV 1759  
Qy 627 LWLAASGLYHS-TNGSSWAITGV-----SSAVNVGFGKAPGSSYPVFWIGGVT 681  
Db 1760 LTDGNGNPICGTLADGSGNWSFTPTPLANGTVNNAVAQDPAGNTSGPASTVDTVAPAT 1819  
Qy 682 GAYESDDCGTWWLINDOHQYGNWQAITGDHANLRVYITGNCRGIVYDGGAPSG 740  
Db 1820 -----PVINPSN-----GSVITGT-AEVGAKVILTDGNGNPICGETTADGSG 1859  
RESULT 12  
ABU19529  
ID ABU19529 standard; protein; 2435 AA.  
XX AC ABU19529;  
XX XX  
XX 19-JUN-2003 (first entry)  
XX Protein encoded by Prokaryotic essential gene #5056.  
XX DE Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX KW Borrelia cepacia.  
XX OS WO200277183-A2.  
XX FN 03-OCT-2002.  
XX PD 21-MAR-2002; 2002WO-US009107.  
XX PF 21-MAR-2001; 2001US-00815242.  
XX PR 06-SEP-2001; 2001US-00948993.  
XX PR 25-OCT-2001; 2001US-0342923P.  
XX PR 08-FEB-2002; 2002US-00072851.  
XX PR 06-MAR-2002; 2002US-0362699P.  
XX XX  
XX FA (ELIT-) ELITRA PHARM INC.  
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX WPI; 2003-029926/02.  
XX DR N-PSDB; ACA23399.  
XX XX  
XX PT New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX XX  
XX PS Claim 25; SEQ ID NO 47453; 1766pp; English.  
XX XX  
XX CC The invention relates to an isolated nucleic acid comprising any one of  
CC of the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing of the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC on a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a

CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by this  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences  
XX Sequence 2435 AA;  
Query Match 4.9%; Score 199.5; DB 6; Length 2435;  
Best Local Similarity 23.0%; Pred. No. 0.00022;  
Matches 185; Conservative 72; Mismatches 303; Indels 243; Gaps 42;  
Qy 47 ANGRW-----IPLLDWGNWNGVVSIAADPI--NTNKVMA----- 83  
Db 173 ANGRWSYTPGTLPFGAHQITVTQDSSGNTSELSTVGIIVDTIVQKTPATAVSDDAQHP 232  
Qy 84 -AVGMYTNSWDPN-----DGAAIL---RSSDQGATWQITPLPKLG----- 119  
Db 233 VTIGGATSDTTPLGCTAEAGSRVSVYDGTLLGTTTADSSGKWTFTPTT-GLGEGAHSI 291  
Qy 120 -----GNMPGRGMGERLAVDPNDNILYFGAPSGKGLWRDTSAGATWSQMTN---FP 168  
Db 292 TVTATDAAGNVSTPSAAFELTITDTPALFTVNAATDGTSLSGTAEAGATVNIIDTGGTP 351  
Qy 169 DVGTVIANPTDTGYQSD---IQGVVWV-APEKSSSLGQASKTIFGVADPNPNVFW 223  
Db 352 D-ATVTADPSGVTWYTPSTPLPIGTIVGTATDAAGNTGPSASVTVGTDTAPGAPVI-- 408  
Qy 224 RDGGATWQAVPGAPTGFIPKHGVFDPVNHVLIATSNCTGCP---YDGS-----SG 270  
Db 409 --GTVTDDA--GSVVGAIASGGSTDDATPTL-SGTAEAGSTVSVYDGTLLGTTTADPSG 463  
Qy 271 DVWKFVSVTS--TWTRISVPSTDTANDYFGVS---GLTIDROHPNTIMV-ATQISWPD 324  
Db 464 N-WTFTPTTGLEGAHSLTATATAGNVSVSTAPDFTLITDTPASIPVTNAID-----G 517  
Qy 325 TIIIFSTDCGATWTRIMWTSYPNRSLRVYLIDISAPPLWTFGVQPNPPVSPKLGWDEA 384  
Db 518 TSLSGTAEAGATINIDTNGDGTDPAT-----VTADPGGAWTVTPSTPLFA----- 562  
Qy 385 MAIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHAPMVKGLEATAVNDLIPSPGAPL 444  
Db 563 -----GTIVIGVAT-----DAAGNTGPSASV-----TVGTDTAP--CAPV 596  
Qy 445 ISALGDLGGFTHADVTAVPSTIFTSPVFTTGSVDYAEINPSIIVRAG-----SFPDS 497  
Db 597 ICTVTDGAGSVVGAISAGSGSTDDATPTL--SGT-----AEAGSTVSVYDGTLLGTTTADPS 651  
Qy 498 SQPNDRHVAFTDGGKWNF-----QOSEPGVT---TGGTVAASADGSRFVWAGDP 546  
Db 652 G-----NMTFTPTTALGEGAHSITVTATDTAGNVSVSTAPDFTLITDTPAP 696  
Qy 547 GQPVVAV--GFCNSWAASGQVPANAQIRSDRVNPKTFYALSNGTFFYRSTDGGVTFQPVAA 605  
Db 697 AIPVTNATDGTSLSGTAEAGATVNIIDTNGDGTDPATVTDPSGAW-----TYTP-ST 747  
Qy 606 GLPSSGAVGVMFHAPVKGEDLWLAASSGLYHSTNGSSWSAITGVSSAYNVGFGKAPG 665  
Db 748 PLPAGTVIGVATDAAGN-----TGPSASVTVGTDTTAPG 782  
Qy 666 SSYPVAVFVGT-----IGGVGTAYRS-----DDCGTTWVLINDOHQYGNWQAITGDHANL 717  
Db 783 AP-----VIGTVTDVGVSVGAIIISGGSTDDATPTLS-----GTAEAGSTVS- 824

QY 718 RRVVIGTNGRGIVYDIGGAPSG 740  
 Db 825 --VVDGTTLLGTTAD----PSG 841

RESULT 13  
 ADA34216  
 ID ADA34216 standard; protein; 3892 AA.  
 XX  
 AC ADA34216;  
 XX  
 DT 20-NOV-2003 (first entry)  
 XX  
 DE Acinetobacter baumannii protein #1377.  
 XX  
 KW Acinetobacter baumannii; bacterial disease; antibacterial; vaccine;  
 KW plant biocontrol agent.  
 XX  
 OS Acinetobacter baumannii.  
 XX  
 PN US6562958-B1.  
 XX  
 PD 13-MAY-2003.  
 XX  
 PF 04-JUN-1999; 99US-00328352.  
 XX  
 PR 09-JUN-1998; 98US-0088701P.  
 XX  
 PA (GENO-) GENOME THERAPEUTICS CORP.  
 XX  
 PI Breton G, Bush D;  
 XX  
 DR WPI; 2003-576092/54.  
 DR N-PSDB; ADA30090.  
 XX

New Acinetobacter baumannii proteins and nucleic acids, useful as reagents for diagnosing a bacterial disease, as components of antibacterial vaccines, as targets for antibacterial drugs, or as biocontrol agents for plants.

Example; SEQ ID NO 5503; 328pp; English.

The invention relates to isolated Acinetobacter baumannii nucleic acids. The A. baumannii nucleic acids and polypeptides are useful as reagents for diagnosing a bacterial disease, as components of antibacterial vaccines, as targets for antibacterial drugs, to detect the presence of A. baumannii and other Acinetobacter species in a sample, in screening compounds for the ability to interfere with the A. baumannii life cycle or to inhibit A. baumannii infection, and as biocontrol agents for plants. The present sequence represents the amino acid sequence of an A. baumannii protein.

XX  
 SQ Sequence 3892 AA;

Query Match 4.5%; Score 181.5; DB 6; Length 3892;  
 Best Local Similarity 21.5%; Pred. No. 0.0072;  
 Matches 204; Conservative 90; Mismatches 344; Indels 309; Gaps 54;

QY 7 TWSNVAICGGGFFVDG-----IVFNEGAPGILYVR-TDIGMYRWDAANGRIWPLLDW 57  
 Db 543 TWT-VSVPGLVADAKTIDAKVTFDDAAGNSSTVNDTQI---YTLDTA-APAAPVIDP 597

QY 58 VGMNNGYNGV--VSIADPINTNKVWAAVGMVYN-----SWD-PNDGAILRS 102  
 Db 598 V-----NGTDRITGTAEPSGTVTV-----TYPNGDTATVVAGPDGSGWSVPNPG--LND 643

QY 103 SDQATWQITPLPKLGN--MPCRGMGERLAVDPNN-----DNILYFGAPSG 148  
 Db 644 GDEVEAIATDP-----AGNPSLPGTATVD--AVGNTDGNVFTVDSVTADNVINASASG 696

QY 149 K----GLWRSTDGATWSQMTNFPDVGTYIANPTDTTGYQSDIQGVVWVAFDKSSSSLGQ 204  
 Db 697 NVTYGV/LKNVPADAAANTVTVINGQTYTATVDSTAG-----TWTVSVPGSDLTAD 748

QY 205 ASKTI-----FVGVADPNPNVFW 222  
 Db 749 ADKTIIDAKVTFTDAAGNSSVNDTQTYTIDTTAPDAPVINPVNGTDPITGTAEPSGTVTV 808

QY 223 SRDGGATWQAVPGAP-TGFI PHKGVPDPVNHVLYTATSTNGPYDGSDDGVKFSVTSCT 281  
 Db 809 TYPDGSSTTVVAGPDGTWTVNPGNLNDG-DKVTATATDPAGNP-----SLPGT 855

QY 282 WTRISPVSTDTANDYFGYSLTIDR-----QHPNTIMVATQI 319  
 Db 856 ATVDAVGENTGVN--FTVDSVTADNVINASASGNVTVTVGLKNVPADAAANTV-VTVVI 912

QY 320 SWPFDTIIFRSTDCGATWTRIMWTSYPNRSRYVLDISAEPMLTF----- 365  
 Db 913 NGQTVTATVDSTAG--TWT-----VSVPGSLDTADADKTIDAKVTFDDAAGNSSVNDTH 965

QY 366 -----GVQPNPVPSPKLGWDEAMAI DPFNSD-----RMLY--GTGATLYATND--- 408  
 Db 966 TTVTVDTVAPNAPV-----LDPINATDPVSGQAEFGSTVTVTPDGTATVVAGPDGSW 1018

QY 409 -----LTKWDSGGQIHI-----APMVKGLEETA VND-----L 435  
 Db 1019 SVPNPGNLVGDGTVTATATDPAGNTSLPGTGTVSADITAPVV-ALDDVLNTDSTPALTGT 1077

QY 436 ISPPSGAPLISA-----LGDLGGFTHAD-----VTAVPSTIFTSPVFTTG----- 475  
 Db 1078 VNDPTATVVVNVGDVYPAVNVNGD-GTWLADNTLPLTDGPHITITVTDAAAGNAGTDT 1136

QY 476 --TSVDYAEALNPSIIVRAGSFPDPSQPNDRHVAFTDGGKNWFGQSEPGGVTTGTVAAAS 533  
 Db 1137 AVVTIDTTAPNAPVLDPINATDP-----VSGTAEAGST-VTVTYPDGT--ATVAG 1185

QY 534 ADGSRFVWAPGDGPQPV-----VYAVFGNSWAASQGVANAOIRSDRVNPKTYFVLSNGT 589  
 Db 1186 TDGS---WSVPNPGNLVGDGTVTAT--ATDPAGNTSLPGTGTVSADITAP--VVALDDVL 1238

QY 590 FVRSDTG--GVTFQVAAAGLPSGAGVGMFHAVPGKEGDLWLAASSGLYHSTHSGSSWSA 647  
 Db 1239 TNDSTPALTGTVNDPTATVVN--VDGTDYPAVNVNGDG-TWTLADNTLPLVADGPHI-IT 1294

QY 648 ITGVSSAVNVG-----FGKSAPGSSYPVAFVWGTIGGTGAYRSDDC-----GTTW 693  
 Db 1295 VTATDAAGNAGTDAVVTIDTTAPNA--PVLDPINATDPVSGTAEAGSTVTVTPDGTTA 1352

QY 694 VLI--NDQHQYGNWGOAITGDHANLRVYIGTN-----GRGIVYGD I 734  
 Db 1353 TVVAGTDGSGWSVPNPGNLVGDGTVTATATDPAGNTSLPGTGTVSADI 1399

RESULT 14  
 ABU16642  
 ID ABU16642 standard; protein; 1308 AA.  
 XX  
 AC ABU16642;  
 XX  
 DT 19-JUN-2003 (first entry)  
 XX  
 DE Protein encoded by Prokaryotic essential gene #2169.  
 KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 XX  
 OS Acinetobacter baumannii.  
 PN WO200277183-A2.  
 XX  
 PD 03-OCT-2002.  
 XX  
 PF 21-MAR-2002; 2002WO-US0009107.  
 XX  
 PR 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX (ELIT-) ELITRA PHARM INC.  
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX WPI; 2003-029926/02.  
DR N-PSDB; ACA20512.  
XX New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX Claim 25; SEQ ID NO 44566; 1766pp; English.  
XX The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway; (8)  
CC required for proliferation, or that inhibits cellular proliferation; (9)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 1308 AA;  
Query Match 4.5%; Score 181; DB 6; Length 1308;  
Best Local Similarity 24.1%; Pred. No. 0.0018;  
Matches 162; Conservative 65; Mismatches 245; Indels 200; Gaps 38;  
Qy 150 GLWRSTDSGATWSQMTNPPDVGTVIANPTDRTGYQSDIQGVVWVAFDKSSSLGQASKTI 209  
Db 2 GSWSVNPG-----NLVGDGTATATADPAG-----NTSLPGTGVSV--- 38  
Qy 210 FVGVDNPNVFWSRDGGATWQAVPGAPTGI--PHKGVPDPVNHVLYATFNTGGPYDG 267  
Db 39 ----ADITAPVV-ALDDVLTLNDSTP-ALTGTVNDPTATVWVNDGVDYPAVNN----- 85  
Qy 268 SSGDVWKFVSTGWTWRI-----SPVSTDTANDYFGYSG-----LTIDRQHNTI 313  
Db 86 --GD-----GIWTLADNLTPLADGPHITVTATDAAGNVGNDTAVVTDITVAPNAP 135  
Qy 314 MVAQTISWMPDTI-----IFRSTDGGATWRIWDWTSYPNRSRLVLDISAEPMLTFGVQ 368  
Db 136 VL-----DPINATDPVSCAQPGSTVT-----VTYPDGTATVAGTGSW----- 176  
Qy 369 PNPVPSPKLV---GWMDEAMAIDPNFSRMLYGTG-----ATLYATNDLTKWDS----- 414

Db 177 ---SVNPNLVDGDTVTATATDP-AGNTSLPOTGTGTVSADITAPVVALDDVLNDSTPAL 232  
Qy 415 GGOIH-----IAPVMKLEETAVNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTS 469  
Db 233 TGTVNDPTATVVVNDGVDYPAVNN-----GD-GTWTLADNLT--PTLADG 275  
Qy 470 PVFTTGTSDVYAEI--NPSIIVRAGSFDSSQPNDRHVAFSTDGKKNWFGSGSPGGVT-- 525  
Db 276 PHTITVTATDAAGNVGNDTAVVTDITVAPNAPVLDPINATDPVSGQ-----ABPGSTVTV 330  
Qy 526 -----TGGTVAASADGSRFVWAPGDPQPV-----VYAVFGNSWAAASQGVPAQAIRSDR 576  
Db 331 TYPDGTATVAGTGS-----WSVPNPNLVDGDTVTAT--ATDPAGNTSLPGTGTVSADI 385  
Qy 577 VNPKTFYALSNGTFYRSTDG--GVTFQPVAAAGLPPSSGAVGMFHAVPGEGDLWLAASSG 634  
Db 386 TAP--VVALDDVLTLNDSTPALTGTVNDPTATVVVN--VDGVDYPAVNNGDG-TWTLADNT 440  
Qy 635 LVHSTNGSSWGAITGVSSAVNVGFKSAPGSSYPVAVVGTGTGGVTVGAYRSDC----- 689  
Db 441 LPTLADGPHIT-ITVTATDAAGNVGVD-----TAVVTIDTVAPKCTGAXSDQCNRLQX 491  
Qy 690 -----GTTWVLINDDHOYGNW-----GQAITGDHANLRVVGITNGR- 727  
Db 492 AVRLEPGSTVTVVTPDGTATVADTDP-----GSWSPNPNLVDGDTCDLSQSYXLSGNT 548  
Qy 728 -----GIVVGDI 734  
Db 549 SLPGTGTVSADI 560  
RESULT 15  
ABU22692  
ID ABU22692 standard; protein; 1074 AA.  
XX AC ABU22692;  
XX DT 19-JUN-2003 (first entry)  
XX DE Protein encoded by Prokaryotic essential gene #8219.  
XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX OS Burkholderia mallei.  
XX PN WO200277183-A2.  
XX PD 03-OCT-2002.  
XX PF 21-MAR-2002; 2002WO-US009107.  
XX PR 21-MAR-2001; 2001US-00815242.  
XX PR 06-SEP-2001; 2001US-00948993.  
XX PR 25-OCT-2001; 2001US-0342923P.  
XX PR 08-FEB-2002; 2002US-00072851.  
XX PR 06-MAR-2002; 2002US-0362699P.  
XX PA (ELIT-) ELITRA PHARM INC.  
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX WPI; 2003-029926/02.  
DR N-PSDB; ACA26562.  
XX New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX Claim 25; SEQ ID NO 50616; 1766pp; English.  
XX The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression



of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide(s) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

xx SQ Sequence 1074 AA;

Query Match 4.2%; Score 170; DB 6; Length 1074;  
Best Local Similarity 21.6%; Pred. No. 0.008;  
Matches 186; Conservative 82; Mismatches 300; Indels 292; Gaps 44;  
  
Qy 10 NVAIG--GGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRIWIPLLDWMVNNWYNG 67  
Db 180 NVGLGQQAGQFVTGT--GNSAMGHLAGST-VSGSY--NAAFG-----EYAGTNTSG-- 225  
  
Qy 68 VVSTADPINTNKVAAVGMVYNSW--DPDNGAILRSSD---QGATWQITPLPFLKGGNMP 123  
Db 226 -----GAN---AAGFGYAGRYINGTNTALGAYDLPVYNGTWYGS---YVTGSNNL 270  
  
Qy 124 GR-----CWGERLAVDPNNNLYFPGSPGKGLWSTDSGATWSQ---MTN 166  
Db 271 GAGHNSGAYVSGASNVGLGDGAGFTVGSNNVAIGTAGSGAYTSGSPGATLNAAALVASN 330  
  
Qy 167 FPDVGTIANPTDTTGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNNPVFWSRDG 226  
Db 331 TVSIG-----TRATASQSDA-----IAGKGATASGAQSISI----- 362  
  
Qy 227 GATWQAVPGAPTGFIPHKGVDPVNNVLYIATSNLTGGPYD--GSSGDYWKFSVTSGTWTRI 285  
Db 363 -GTGNVSVSGKSG---AIGDP-----STVSGAGSYSIGNNTV---ANSNTFVLG 405  
  
Qy 286 SPVPSTDTANDYFGYSLGTLTDQHPNPTIMVATQISWHPDTHIFRSTGGATWTRWDWTS 345  
Db 406 NGVTTT-----QDNSVVLGNQSTDRAAVAVSSETINGTTY-----N 441  
  
Qy 346 YPNRSLRYLVDISAEPWLTFCVQNPVPVPPKLGMDMAIDPPFNDRMLYG----- 398  
Db 442 YAG-----VASPANG-----VVSIGGVGTERQLINVAAGQVS 473  
  
Qy 399 -----TGATLYATNDLTKWDSGGQIHIAPMVKGLEETAVNDL----- 435  
Db 474 ATSTDAINGSQLYATN-----QAVIA-----EDAKVNSLGGGVASALGGNAAYNA 518  
  
Qy 436 -----ISPPSGAPLISALGDLGGFTHADVAVTPSTIFTSVPTTGTSDVYAEINPSIIVRA 491  
Db 519 TTGATITAPSYAVVGTQNSVGGAIQALQALPQYTSVGPVTTTPNAPGSAPTNTVTLVGA 578  
  
Qy 492 GSPDPSSQP-----NDRHVAFFSTD-----GCKNWFQSGSEP 521

Db 579 GPGANITTPVTLTNAVPGKLSATSTDAVNGSQLYATNQVANLVSNNGGVGFVQYSDP 638  
Qy 522 GGVTT--GG-----TVAASADGSRFVWAPGDPGQPVVYAVFGNSWAAASQGVYAPNAQIR 573  
Db 639 SAPITPNGGKPSQDLTLVGAASGPVALHNVA--PGTASTDAVNVGQLGAVTTGLGGGAAI- 696  
Qy 574 SDRVNPKT-----FYALSN--GTFYRSTDGGVTTQPV-AAGL-----PSSGAVG- 614  
Db 697 ----DPKTAGVTPSYTYVNADGTTNSVNVGAAIDAINSTGIKYFHANSTKPDQALGA 752  
Qy 615 ----VMPHAVPKEGDLWLAASSGLYHSTNGSSWSAITGVS-----SAVNVGFGK 661  
Db 753 DSAVIGPNAVANNAGD--VALGSGAVTSOAGTSLSETINGVTYSFAGTTPIGTVSVG--- 807  
Qy 662 SAPGSSYPVAFVVGTVGTGAYRSDDCGTTWVLINDDQHQYGNWG-QAITGDHANLRRV 720  
Db 808 -APGVERT---ITNVAAGRIQSSSTD-----ALNGSOLYGTNQSI EALTDKNSLGNT 856  
Qy 721 YITNGRGIVYDGIIGGAPSG 740  
Db 857 VANTLIGSGASYNPQTGA VNG 876

Search completed: May 11, 2004, 12:06:49  
Job time : 125.24 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 11:51:45 ; Search time 34.7001 Seconds  
(without alignments)  
2051.340 Million cell updates/sec

Title: US-09-917-376-3  
Perfect score: 4036  
Sequence: 1 ATTQPTYWNSVAIGGGFVD.....YIGTNGRGIVYDIGGAPSG 740

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR.78.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	2371	58.7	890	2 T35237	probable secreted
2	2009	49.8	839	2 D97033	probably secreted
3	1680	41.6	856	2 T00349	Avicelase III - As
4	1160.5	28.8	707	2 F72393	hypothetical prote
5	238	5.9	2468	2 A83412	hypothetical glyci
6	202.5	5.0	2174	2 E95965	hypothetical prote
7	189	4.7	3972	2 S75231	hypothetical prote
8	186.5	4.6	4199	2 S76412	hypothetical prote
9	182	4.5	1904	2 T13256	tail-host specific
10	175.5	4.3	2523	2 F70846	probable PPE prote
11	172	4.3	993	2 AE1905	outer membrane sec
12	171	4.2	908	2 AE2254	hypothetical prote
13	170.5	4.2	5188	2 B85547	probable RTX fami
14	168.5	4.2	5291	2 F90696	hypothetical prote
15	168	4.2	2124	2 A28452	proteoglycan core
16	167.5	4.2	2554	2 AB3528	extracellular seri
17	166.5	4.1	902	2 H87323	hypothetical prote
18	166.5	4.1	980	2 H90681	probable flagellin
19	166	4.1	699	2 D70533	hypothetical prote
20	166	4.1	1468	2 A44345	nucleoporin - rat
21	165.5	4.1	980	2 D85532	probable structura
22	163.5	4.1	618	2 T49741	related to stress
23	163.5	4.1	13055	2 T16580	hypothetical prote
24	163	4.0	1684	2 S10789	anyase A-180 - al
25	160.5	4.0	3716	2 E70969	probable PPE prote
26	160	4.0	1341	2 H98323	hypothetical prote
27	160	4.0	3624	2 AD0835	large repetitive p
28	158.5	3.9	1441	2 B86807	hypothetical prote
29	158	3.9	2232	2 T34434	hypothetical prote

30	157	3.9	809	2 A55547	quinase-shikimate
31	156.5	3.9	2204	2 A70524	probable PPE prote
32	156	3.9	3157	2 B70969	probable PPE prote
33	155	3.8	4180	2 G83559	hypothetical prote
34	154.5	3.8	1032	2 T34433	hypothetical prote
35	154.5	3.8	3570	2 T45025	mucin MUC5B, trach
36	153	3.8	1055	2 A87364	OMP-related prote
37	153	3.8	1821	2 AG2335	hypothetical prote
38	153	3.8	13288	2 T03099	mucin, submaxillar
39	152.5	3.8	3016	2 S77300	hypothetical prote
40	152	3.8	348	2 T35248	probable oxidoredu
41	152	3.8	1049	2 T42045	beta transducin-li
42	152	3.8	1196	2 A29130	beta-amylase (EC 3
43	152	3.8	3472	2 T31308	hypothetical 367K
44	151.5	3.8	820	2 B72575	hypothetical prote
45	151.5	3.8	1145	2 B75625	hypothetical prote

ALIGNMENTS

RESULT 1

T35237

probable secreted cellulase - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor

C>Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 05-Nov-1999

C:Accession: T35237

R:Seeger, K.J.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, September 1998

A:Reference number: Z21572

A:Accession: T35237

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-890 <SEE>

A:Cross-references: EMBL:AL031515; PIDN:CAA20642.1; GSPDB:GN00070; SCOEDB:SC5C7.30C

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCOEDB:SC5C7.30C

Query Match 58.7%; Score 2371; DB 2; Length 890;

Best Local Similarity 57.7%; Pred. No. 1.4e-131;

Matches 431; Conservative 109; Mismatches 183; Indels 24; Gaps 10;

QY	6	YTWNSVAIGGGFVDGIVFNEGAPGILYVRTDIGYRWDAANGRWIPLLDWVGNWGY	65
DB	41	YTWKNARIDGGGFVGVFNTEKDLAYARTDIGAYRQWESHTWTPLLDHVGWDDWGH	100
QY	66	NGVYSIAADPINTNKKVAAVGYMNSDPNDGAILRSSDQGATWQITPLPFLKLGNNMGR	125
DB	101	TGVVALASDAVDPRVYAAVGYTYNDWPTNGAVLSADRGASWEKADLPFLKLGNNMGR	160
QY	126	GMGERLAVDPNNDLILYFGAPSGKGLWRSTDSGATWSQMTNFPDVGTIYANTDTTGYOS	185
DB	161	GMGERLAVDPNDVLYLGAPSGHGLWRSTDSGATWSQMTNFPDVGTIYANTDTTGYAS	220
QY	186	DIQGVWVAVDPKSS--SSLGQASKTIFVGVDPPNNVFWSRDGGATWQAVPGAPTGFPHK	244
DB	221	DNQGITWTFDESGGGAGTARTIUVGVADKENVYRSTDAWTERLAGQPTGYLAHK	280
QY	245	GVFPVPHVLIATISNTGGPYDSSGDVWKFSTGTTWTRISPFVSTDTANDYFYGSGLT	304
DB	281	GVLDANGYLYLAYSDDTGGPYDGGKGLRYATATGTWTDISPAEADI---YFGSGLT	337
QY	305	IDRQHPNTIMVATQISWWPDTIIFRSTGGATWTRIDWTSYPNLSRLVRLDISAPWLT	364
DB	338	VDRQRPQVWATAYSSWWPDTQIFRSTDSGATWSQAWSTSYSPDRNRYTMDVSSSPWLT	397
QY	365	FGVQNPVPSPKLGWDEMAIDPFNSDRMLYGTGATLYANDLTWK--DSGQIHPM	423
DB	398	WGAMPAPPEQTPKLGWMTAELEIDPFSDRMVYGTGATYGTENTLNWDEGGTFAVEPM	457
QY	424	VKGLEETA VNDLISPPSGAPLISALGDLGGFTHADVAVPSTIFTSPVTTGTSVDYAEI	483

Db 458 VRGLEETAVNDLASPPSGAPLLSALGDVGGFRHTSLTEVPSSMYTSPNFTSTSLDPAET 517

QY 484 NPSIIIVRAGSPDPSSQPNDRHVAFTDGGKWNFGQSGEGVTTGCTVAASADGSRFWAP 543

Db 518 KPDVVVRAGNLD--SGP---HIAFTDNGANWFGGTDPSGVSGGCTVAAGADGSRFWSP 572

QY 544 GDPQPVVYAVGFGNSWAASQGVPAQAIRSDRVNPKTFYALNSGTFFRSTDDGGVTFQPV 603

Db 573 --EGAGVQVTTGFGTSWAQSTGLPAGAVESDRVNPATFYGFKSGRFYVSTDDGATFTAS 630

QY 604 AA-GLPSSGAVGVMTFHPVKEGDLWLAASS-----GLVHSTNGGSSWAITGVSSAVNV 657

Db 631 AATGLPAGD--GVRFKALPGEEDVWLAGGAADGYGLWHSTDDGGCTFRLFEVDAADTV 688

QY 658 GFGKSAPGSSYPVAVFVGTIGVGTAYRSDDCGTTWVLINDDHOYGNWGAITGDHANL 717

Db 689 GFGKAAPGASYQLTFTSAEIGVGIRFSTDDAGATWTRVNDDAHQMGWGTGAITGDPRVY 748

QY 718 RRVYIGTNGRGIVYGDY----GGAPSG 740

Db 749 GRVYVATNGRGIVYGDYSDTGGGTDPG 775

RESULT 2

D37013

probably secreted sialidase, several ASP-boxes and dockerin domain [imported] - Clostrid

C:Species: Clostridium acetobutylicum

C>Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001

C:Accession: D37013

R.Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo

A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: D37013

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-839 <KUR>

A:CROSS-references: GB:AE001437; PIDN:AAK78895.1; PID:GI5023820; GSPDB:GN00168

A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:

A:Gene: CAC0919

Query Match 49.8%; Score 2009; DB 2; Length 839;

Best Local Similarity 50.9%; Pred. No. 2.4e-110;

Matches 377; Conservative 107; Mismatches 240; Indels 16; Gaps 12;

QY 3 TQPYTWSNVAICGGGFVDGIVNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWGNW 62

Db 37 SQGYKWDNAKIGAGGYVPAVFNKTEKDLIYARTDMGAYRWKANNKWIPTD--GFSD 94

QY 63 WGYNGVVSIAADPINTNKVMAAVGMVMTNSWDNDGAILRSSDOGATWQITPLPKLGGNN 122

Db 95 WTMJGCESIATDPIDITNKVYIAAGLYTNDMDENAYILSSQDKGNTWKRYQPLPKVGGNN 154

QY 123 PGRGMGERLAVDPNNNDILYFAPSGKGLWRSTDSGATWSQMTNPPDVGTYIANPTDITG 182

Db 155 PGRNMGRLQIDPNDKILYLGARSNGLWKSSEDYGTWSKVDFNPDGTYVDQONE-- 212

QY 183 YOSDIQGVVWVAFDKSSSLGQASKTIIFGVAD--PNNPVFWSRDCGATWQAVPGAPTGI 241

Db 213 YTADKGVVWVETFDPSGTGKSPSTQTMVGAADTKGNNIYVTDGDKTWSAVKGPQKYL 272

QY 242 PHKGVFDPVNHVLIYATNTGPGYDSSGDVWKFVSTSGTWTRISPVSTDTANDYFGVS 301

Db 273 PHHGIL--ASDGLMLYISYNTGPGYDSSGDVWKFVSTSGTWTRISPVSTDTANDYFGVS 328

QY 302 GLTIDRQPNPTIMVATQISWNPDTIIFRSTDCGATWTRIDWNTSPNRSRLRVLDISAEP 361

Db 329 GISVDAQPNVNVVATLNRPWDFDEIYRSTDAKTKWKFINDWNGVFNRTILGNVLDYSAQP 388

QY 362 WLTFG--VQPNPVPVFPKLGWMDMAIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420

Db 389 WLDWCKTGTVTPDDPLVKLGMMGDLIEDFPNSDRMFYGTGATLYGTDDLTNWDGKNVDI 448

QY 421 APMVKLEETAVNDLISPPSGAPLLSALGDLCGFTHADVTAVBSTIFTSPVFTTGISVDY 480

Db 449 SVKANGIECAVNDVVVPTKGAQLLSAVGDCGCFYHDDITKVPKMTTPNFSATTSIDY 508

QY 481 AELNPSIIIVRAGSPDPSSQPNDRHVAFTDGGKWNFGQ--GSEPGGVTTGGTVAASADGSRF 539

Db 509 AESSVFNVVRVGVNDVTSKNQDKCGISYDGGKWNFSAGSNISGVYKAGTVAAGADAKTI 568

QY 540 VMAPGDGQPVVYAVGFGNSWAASQGVPAQAIRSDRVNPKTFYALNSGTFFYESTDGGVT 599

Db 569 WSP--EEGANAAYSTDNKNTKTPCSGLPQCAKVRSDRVNPKTFYGFINGKFIYSTDAGAT 627

QY 600 F-QPVAAGLPPSSGAVGVMTFHPVKEGDLWLA--ASSGLYHSTNGGSSWAITGVSSAVNV 657

Db 628 FTQSSQTGLPTKKG-GI-FKTVIGHGEDIWAGKDLWHSTDSGATFTKVGSDVADSDTV 685

QY 658 GFGKSAPGSSYPVAVFVGTIGVGTAYRSDDCGTTWVLINDDHOYGNWGAITGDHANL 717

Db 686 GLGKSKTDDGYPAIYMDATIDGTAGIFRSDDEGATWTRVNDDAHQYGSYDCTGDPNKY 745

QY 718 RRVYIGTNGRGIVYGDIGGA 737

Db 746 GRVFGTNGRGIVYGDIDGS 765

RESULT 3

T00349

Avicelase III - Aspergillus aculeatus

C:Species: Aspergillus aculeatus

C>Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 16-Jul-1999

C:Accession: T00349

R.Arai, M.; Takada, G.; Kawaguchi, T.; Sumitani, J.

submitted to the EMBL Data Library, June 1998

A:Description: Avicelase III from Aspergillus aculeatus.

A:Reference number: Z14141

A:Accession: T00349

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-856 <ARA>

A:CROSS-references: EMBL:AB015511; NID:d1199887; PID:d1029971

C:Genetics:

A:Gene: aviii

C:Superfamily: fungal cellulose-binding domain homology

F:823-854/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 41.6%; Score 1680; DB 2; Length 856;

Best Local Similarity 46.2%; Pred. No. 5e-91;

Matches 346; Conservative 113; Mismatches 258; Indels 32; Gaps 17;

QY 1 ATTOPYTWNSVAI-CGGGFVDGIVNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVG 59

Db 21 AASQAYTWKQNVVTGGGGFTPGIVFNPSAKVAYARTDIGGAYRLN--SDDTWTFPLMDWVG 79

QY 60 ---WNNWGVYVVSIAADPINTNKVMAAVGMVMTNSWDNDGAILRSSDOGATWQITPLPF 116

Db 80 NDTWHDW---GIDALATDPVDTRVYVAVGMVMTNEWDPNVGSILRSTDQDGTWTEKLPF 136

QY 117 KLGGNMPGRGMGERLAVDPNNNDILYFAPSGKGLWRSTDSGATWSQMTNPPDVGTYIAN 176

Db 137 KVGGMNPGRMGERLAVDPNKNISILYFGARSGHGLWKSTDYGATWSNVTSFTWGTGYFQD 196

QY 177 PDITTYGQSDIOGVVWVAEDKSSSLGQASKTIIFGVADPNNPVFWSRDCGATWQAVPGA 236

Db 197 SSST--YTSDDPGIAWVTFDSTSGSSGSAIPRIFGVADAGKSVFKSEADAGATWAWVSGE 254

QY 237 PT-GFIPHKGVDPVNVHVIYATNTGPGYDSSGDVWKFVSTSGTWTRISPVSTDTAN 295

Db 255 PQYGLPHKGVLSPEKTLIYSTANGAGPYDGTNGTVHKNITSGVWTDISP---TSLAS 311

QY 296 DYFGYSGLTIDRQHPNTIMVATQISWNPDTIIFRSTDCGATWTRIDWNTSPNRSRLRYVL 355

Db 312 TYGYGGLSVDLQVPGTLMVAALNCWPPDELI FRSTDSGATWSPIEWNGVPSINYYSY 371  
Qy 356 DISAEPWLTGCVOPNP-PVPSPKLGMDEMAIDPFNSDRMLYCTGATLATNDLTKWDS 414  
Db 372 DISNAPWIDTSTSDQFPV---RVGMVVEALADPFDSNHWLYCTGLTVYGGHDLTWDS 428  
Qy 415 GGQIHIAPWVKGBETAVNDLISPPSGAPLISALGDLGGFTHADVAVPSTIFTSPTFT 474  
Db 429 KHNVTKSLAVGIBEMAVLGLITPPGFPALLSAGVDDGDFVHSDLDAAPNQAYHTPTYGT 488  
Qy 475 GTSVDVAELNPSIIVRAGSPSPQPNDRHVAFSTDCGKWFQSGFPGVTTGCTVAASA 534  
Db 489 TNGIDYAGNKPNSIIVRSASD--DYPT---LAUSNFGSTWIADYAASTGTCAVALSA 543  
Qy 535 DGSRFVWAPGDPGPVVVAVFGNSWAASQCPANAOIRSDRVNPKTFYALNSGTFYRST 594  
Db 544 DGDIVLLMSTSGALVSKSQG---TLTAVSLSPGAVIADKSDNTVYFGSGSAGAIYVK 600  
Qy 595 DGGVTQFPVAAGLPSSGAVGMFHAVPGKEGDLWLAASSGLYHSTNGGSSSAI-TGVSS 653  
Db 601 NTATSFTKTVS-LGSSTTVNAI-RAHPSIAGDVWASTDKGLWHSTDYGSTFTQIGSGVTA 658  
Qy 654 AVNVGFKSAPGSSYPVAVFVGTIGTGVAYRSDCGTWTWVLINDDOHQYGNWQA--IT 711  
Db 659 GWSFGKASSTGSYVVIYGFFTIDGAAGLFKSBDAGTNQVISEDASHGFGS-GSANVNV 717  
Qy 712 GDHANLRVYGTNGRGIVYGDIGGAPSG 740  
Db 718 GDLQYGRVFRGHRPGHLLRQSQREPAG 746

RESULT 4  
F72393  
hypotheical protein TM0305 - Thermotoga maritima (strain MSB8)  
C:Species: Thermotoga maritima  
C>Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C:Accession: F72393  
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;  
C.M.  
Nature 399, 323-329, 1999  
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq  
A:Reference number: A72200; MUID:99287316; PMID:10360571  
A:Accession: F72393  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-707 <ARN>  
A:Cross-references: GB:AE001712; GB:AE000512; NID:94980799; PIDN:AAD35393.1; PID:g498080  
A:Experimental source: strain MSB8  
C:Genetics:  
A:Gene: TM0305

Query Match 28.8%; Score 1160.5; DB 2; Length 707;  
Best Local Similarity 35.2%; Pred. No. 1.2e-60;  
Matches 264; Conservative 123; Mismatches 275; Indels 89; Gaps 26;

Qy 6 YTSNVAIGGGFVDGIVFNEGAPGILYVVRTDGGMYRWDAAAGRWIPLLDWGVNNGY 65  
Db 21 FEWKSVEINGGFPVGIIFHPASPGLLYARTDVGLYRWDEETKRWKQLDFDLRDSQSDY 80  
Qy 66 NGVYSIAADPINTNKVAAVGMVYNSWDPNDGAILRSDQCATWOITPLP---FKLGNNM 122  
Db 81 MGVLUSVALDFSDPKRIYAMTKYTDW-AGYGALISDYGETWTIIVNLKYGKVGNGNE 139  
Qy 123 PGRGMGERLAVDPNNDNLIYPGAPSGKGLWRSTDSGATWSQMTNFPDPVGTYIANPTDTTG 182  
Db 140 DGRNAGERLQVDPNFSVLFMGT-TKYLGLWKSDEFGKNKKVDSFPST----- 186  
Qy 183 YQSDIQGVWVAVFKSSSSLGQASKTTFVGVADPNNPVFWSRDSGATWQAVPGAPTGP 242  
Db 187 -----SVTFVLDFEKSGEKSGSPTRIFVCGSEPKG-IFVTEDDGTTWNVLNPLNPLIP 239  
Qy 243 HKGVDFPNVHLYIATSNVCGPYDSSGDVWKFVTSQWTRISPVPSPTDANDYFGVSG 302

Db 240 LRKGIH--DGLYVLTLSNALGNGATRGAVMKYVIADQKQWYDVTPMKGD-----FGYCG 291  
Qy 303 LTIDRQHEHTIWAQISWNPDTIIFRSTGCGATWTRIWDWTSYPNRSRLRYVLIDISAEPW 362  
Db 292 IDVQE--NVIVSTLDRWPHDEIFISLNGGETWRPILLEKANF-----DINKAPW 339  
Qy 363 LTFGVQPNPPVPSPKLGWMDAMADPPNDRMLYGTGATLYATNDLTKWDSGGQIHIAP 422  
Db 340 IK--DLNP-----HWISD-VKIDPFDNRRAIFTTGYGVWVYELKKSFEF---MGK 384  
Qy 423 MVK-----GLEETAVNDLISPPSGAPLISALGDLGGFTHADVAVPSTIFTSPTFTGT 476  
Db 385 PVKIFENRGLSEETVVLQVPEIGRPLLSATADMGFRHESLDTDPSSMY-KPLKWTSL 443  
Qy 477 SVDVAELNPSIIVRAGSPSPQPNDRHVAFSTDCGKWFQ-GSEPGGVTTGG--TVAAS 533  
Db 444 GIAFAYQNSKFVARVHTY---TYP---FLSYSEDDGINWREIETVPEGITDGGRLSLAVS 497  
Qy 534 ADGSRFVWAPGDPGPVVVAVFGNSWAASQCPVPA---NAQIRSDRVNPKTFYAL--SNG 588  
Db 498 NDKTLVWSPAN--HEVIVSSDKGSKWKAISVPVPEPNYFPASDPVNFPSKFYIFDWKNG 555  
Qy 589 TFRSTGCGVTQFPVAAGLPSSGAVGMFHA-----VPGKEGDLWLAAS-SGLYHSTNGGS 643  
Db 556 DFLISKDGGKGFMK-GAKLPSFDNWNWVSLYSPVLAPDREGDIWLALQWNGLYRSKDGGI 614  
Qy 644 SWSAITGVSSAVNVVCGFKSAPGSSYPVAVFVGTIGTGVAYRSDCGTWTWVLINDDOHQY 703  
Db 615 TFERLGNVDIAVYIGFGAPKPGTDYPAIYLNQWNGVYGFIMSTDGKTMWRINNDKHQF 674  
Qy 704 GNWGOAITGDHANLRVYGTNGRGIVYGD 734  
Db 675 G-WIHYMLGDMNEFCRIFLTGTEGRGIIVGEV 704

RESULT 5  
A83412  
hypotheical protein PA1874 [imported] - Pseudomonas aeruginosa (strain PAO1)  
C:Species: Pseudomonas aeruginosa  
C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: A83412  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bri  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic patho  
A:Reference number: A82950; MUID:20437337; PMID:10984043  
A:Accession: A83412  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-2468 <STO>  
A:Cross-references: GB:AE004613; GB:AE004091; NID:g9947856; PIDN:AAG05263.1; GSPDB:GN001  
A:Experimental source: strain PAO1  
C:Genetics:  
A:Gene: PA1874

Query Match 5.9%; Score 238; DB 2; Length 2468;  
Best Local Similarity 22.8%; Pred. No. 9.1e-06;  
Matches 205; Conservative 88; Mismatches 314; Indels 292; Gaps 50;

Qy 11 VAIGGGGFVDGIVFNEGAPGILYVVRTDGGMYRWDAAAGRWIPLLDWGVNNGY----- 65  
Db 1084 VNLSSGSLSG---TABPFGSTVIITD-----GNGNPFAETADSGNWTYTPSTP 1130  
Qy 66 --NG-VVSTAADPINTNKVAAVGMVYNS-----WDPNDGAILR--SSQCATWOITPLP 115  
Db 1131 IANGTVNVVQAQDASGNSPPATVTVDSAPPAPVNPNSGVVISCTAERAGATVTLT--- 1187  
Qy 116 FKLGNMFPGRGGERLAVDPNNDNLIYPGAPSGKGLWRSTDSGATWSQMTNFPDVGTYIA 175  
Db 1188 -DAGGNPIQGVTDAD-----GSGNWSFTP-----GTPLA 1214

```
QY 176 N-----PTDTTQYQS---DIQGVVWVAFPKSSSL--QOASKTIFVGVADPN-N 218
Db 1215 NGTVIVATADPTGNTGQAAATVDAVAPPVADIPDSNGTTISGTAAEAGAKVILTDGNGN 1274
QY 219 PV-FWSRDGGATQAVPCAP-----TG-----FIPHKGVDFDPN 251
Db 1275 PIGETTADGSGNWFPTGTPLANGTVNNAVAQDPAGNTGPOGSTTVDAVAPTPVNFPSN 1334
QY 252 HVLVIATSNTOGPY---DGSSGDVWKFST--SGTW--TRISPVFS-----TDTANDYFGYS 301
Db 1335 GNLLNGTAEPGSTVTLTDGNGNPIGOTTADGSGNWFPTGSQLPNGTVNVTASDAAGNT 1394
QY 302 GL-----TIDROHPNTIMV-----ATQISWPD---TIIIFRSTDDG-----ATWTRIDW 342
Db 1395 SLPATTTVSSLPSPQPDPNSGVISGTADAGNTIIII--TDGNGNPIGQVTDAGSGNWS 1452
QY 343 WT-----SYPNRSIRYLVLDISAEPLW---TFGVQPNPPVPSPKLGWDEAMATDPNS 392
Db 1453 FTPGIPPLDGTIVNNAKSPNSVDSPAPVITVDGVAPAPV-----IDPSNG 1499
QY 393 DRMLYGT---GATLYATNDLTKWDSGGQIHIAPMVKGLEETAVNDLISP-----PSGAPLI 445
Db 1500 TE-ISGTAEAGATVILT-----DGGN-----PIQATADGSGNWFPTPSTPLANGTVIN 1548
QY 446 SALDGLGGFTH-----ADTVAPSTIF--TSPVFTTGTSDVYAEIPLNPSIIVRAGSPDS 497
Db 1549 AVAQDPAGNTSGPASVTVDALAPAPVINPNSGVVISGT---AEAGATVILTDCGNPI 1604
QY 498 SOPNDRHVFSTDCGKNWF-----QSGEPGGVT----- 525
Db 1605 GQ-----VTADGSGNWSFTPGTPLANGSVINALAQAAGNNSPTSATVDSLAAPV 1657
QY 526 ---TGTVAAASADGSRFVWAPGPGQPV---VYAVFGNCSWAASQGVFANAIQIRSDRVP 579
Db 1658 IDPSNGSVIAGTAAGATVILTDGNGNPIGQVTDAGSGN-WSFTPGTP----- 1704
QY 580 KTFVALSNGTYR-----STDGGV--TFQPVAAGLP-----SSGAVGMFHAVPGKEGD 626
Db 1705 ---LNGTVNNAVAQAAGNTSGPVSTTVDAVAPATFVIDPSNGVELSGTAEPGRVI 1759
QY 627 LWLAASSGLYHS--TNGSSSWAIGTV-----SSAVNVGKSKAPGSSYPVAVFVGTIGVTT 681
Db 1760 LTDGNGNPIGQTLADGSGNWSFTPGTPLANGTVNNAVAQDPAGNTSGPASTVTDVAPAT 1819
QY 682 GAYRSDDCGTTWVLINDDOHQYGNWGOAITGDHANLRVYIGTNGRGIVYGDIGCAPSG 740
Db 1820 -----PVINPSN-----GSVITGT-REVGAKVILTDGNGNPIGETTADSGS 1859

RESULT 6
E95965
Hypothetical glycine-rich protein [imported] - Sinorhizobium meliloti (strain 1021) mag
C:Species: Sinorhizobium meliloti
C>Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: E95965
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A:Title: The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-fixing endo
A:Reference number: A95642; MUID:21396508; PMID:11481431
A:Accession: E95965
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2174 <KUR>
A:Cross-references: GB:AL591985; PIDN:CAQ49389.1; PID:g15140875; GSPDB:GN00167
A:Experimental source: strain 1021, megaplasmid pSymb
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A96039; MUID:21368234; PMID:11474104
```

A:Contents: annotation  
C:Genetics:  
A:Gene: Smb21548  
A:Genome: plasmid

Query Match 5.0%; Score 202.5; DB 2; Length 2174;  
Best Local Similarity 22.7%; Pred. No. 0.00093;  
Matches 176; Conservative 90; Mismatches 312; Indels 197; Gaps 36;

```
QY 12 AIGGGGFDGIVFNEGAPGILVVRT-DIGMYRWDAAANGRWIPLLDVGMNWNWYNGVWS 70
Db 976 SIGGGG-----GNGGVGSVNSKEIGSFNLTA-----NVGVGSGS 1011
QY 71 IAADPINTKNWAAVGMVMTNSWDPNDGAILRSDOGA-----TWQITPLPFK 117
Db 1012 GCASGCGN---AIVGL-----DSGTHLQTSGSGARGVIVQSIGGGGTSCQASVGLS 1059
QY 118 LGGNWPGRMGRLAVDPNNDNLIYFGAPSGKGLWRSTDSGATWSQMTNPFVGTYYANP 177
Db 1060 ASASLPGGG-EEAAEAEESEEGSAGFSASVSVGRTGGSGS-SGTVNVTTAGT----- 1112
QY 178 TDTTGVQSDIQGVWVAFDKS---SSILGOASKTIFVGVADPNPVPFWSRDG----- 226
Db 1113 --ISTFGADADGVLAQSIGGGGGLGSGVQASGSDSEPLDDEGSECASGNGDDGHGY 1170
QY 227 -----GATWQAVPGAPTGFIPHKGVDFPNVHLYIATS-----NTGGPYDGSSG 270
Db 1171 CFGVSVGAT---IDGGTGTGAANGNAVTLTHAGHATATAGDWADGIVAQSIGG--GGAG 1225
QY 271 DVWKFSVTSGTWTRISPVSTDTANDYFGVSLTIDROHPNTIMVATQISWPDITIFRS 330
Db 1226 GTSTAGSGQATANIITVGVGGSGAGNGGAVGITFDDNHGNSISTAGSAY---GVLLQS 1282
QY 331 TDGGATWTRIDWTSVPNRSRLVLD-----ISAEPLWITFGVQPNPVPFS 375
Db 1283 IGGGGG---QGGDGSDEAAGRITVGGFGGSGGSGGMVTAKGWINLSTSG----- 1332
QY 376 PKLGMWDEAMAI--DPFNSDRMLYGTGATLYA-----TNDLTWDSGGQIHIAPMVKGL 427
Db 1333 -----DDAHGIVAQSIGGGGVGAGSSTAEEKHSHTIDLVVGGSG-----VGGSGGE 1382
QY 428 EFTAVNDLISPPSGAPLISAL-----GDLGGFTHADTVAPSTIPTSTPFTTGTSTVDY 480
Db 1383 VDLSTGTTLS--TSGARALGLVAQSIGGGGIGGAGEAD--SIASLVVGG---SGGTIDG 1436
QY 481 AELNPSIIVRAGSPSPSQPNDRH--VAFSTDGGKN-----WFGSECGGVTTGG 528
Db 1437 GAVTVDLTSQSSI---TTQGIAAHGLVAQSIGGGGVGGAAGAPLSFTGNSPGSGDGG 1493
QY 529 TVAASADGSRFV---WAPGDPGPVVYAVFGNCSWAASQGVFANAIQIRSDRVPKTFYAL 585
Db 1494 DVAVTADGSIITRGDYAFGLVAQSIGGGGFGGN-ATSAFIGSNGNLSSDG----- 1543
QY 586 SNGTFYRSTGGVTFQPVVA-----AGLPSGAGVGMFHV-APVGKEGD---LWL 629
Db 1544 KSGNVTVSLDAGRTQASCKDSIGIFPAQSDAGTDNNGTIDVTVNGTGTGSGDNGAGIIV 1603
QY 630 AASSGLYHSTNGCSWSALITGVSSAVNVGFKSAPGSSYPVAVFVGTI--CGVTGA 683
Db 1604 SAGKNVTVNSGNGVSAASGV--AVQYTAGMNSPEDSTLLVNVNAGTITSGSVKGA 1656
```

## RESULT 7

S75251

Hypothetical protein slr1028 - Synecocystis sp. (strain PCC 6803)

C:Species: Synecocystis sp.

A:Variety: PCC 6803

C&gt;Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000

C:Accession: S75251

R:Kaneke, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;  
O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda,  
DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis

S. Query Match 4.78; Score 189; DB 2; Length 3972;  
Best Local Similarity 19.3%; Pred. No. 0.013;  
Matches 166; Conservative 93; Mismatches 276; Indels 326; Gaps 42;

A;Reference number: S74322; MUID:97061201; PMID:8905231  
A;Accession: S75251  
A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-3972 <KAN>

A;Cross-references: EMBL:D90904; GB:AB001339; NID:g1652225; PIDN:BAAL7165.1; PID:g165224  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C;Superfamily: Synchocystis hypothetical protein slr1028

Query Match 4.78; Score 189; DB 2; Length 3972;  
Best Local Similarity 19.3%; Pred. No. 0.013;  
Matches 166; Conservative 93; Mismatches 276; Indels 326; Gaps 42;

QY 13 IGGGFGVDGVFNEGAPGILYVRTDIGMYRWDAAANG-----RWIPLLDW- 57

DB 2458 VNGDGFADVLISGSPAGGVILFGNSTKDLL--DAALGTDDLIIISVENAQVKEFVALGDFD 2515

QY 58 -----VGMNMGYNG---VVSIAADPINTKVAAYGMVYNSWD-- 93

DB 2516 GDGLADPGVDDQGNFFLVGLSPGLSGSLVIDSTLPNLSNFQAWGVGDENGVDYDF 2575

QY 94 -----PNDGAILRSDQGAQTQITPLPKLGNMP-----GRGMGERLAVDPN-- 136

DB 2576 VLOGPNSTIAVYGNANGTLTDSPLTF--GNFPLPSSFTGIDLNGIKIIEIVAGQPNLN 2633

QY 137 -----NDNILYFGAPSKGLWRST-----DSG-ATWSQMTNFPDVGTYIANP 177

DB 2634 PVPNIGFGGGLQFTYEAGNAVLQPTVPPNASVTEASGLSSWGQI-SFPNQVAQAGVP 2692

QY 178 TDTT--GY-----QSDIQGVVW-----VAPDKSSS----- 200

DB 2693 SFATLDGMLYQAFYGINERISTKDSYIYIQRSDGVSWENLTQVPLDSNGTPIDLKNLP 2752

QY 201 -SLGQASKTIFVGVDAPNNPVFWSRDG-----GATWQAVP-----GAPT--GFTPHK 244

DB 2753 PSITAYNGTLYLGTADNGQV--WVAEGVNTNANSGLINAVPINOASNGNPTLVAFNDEL 2811

QY 245 GVF-----DPNVHLYATNTGGPYDSSGDVWKFSTGTWTRISVPVSTDTANDYEGY 300

DB 2812 YVFFVKDASNDILYSSSNPG-----SSSG--WDGTSTVLTFSDVNOATN-----FPL 2858

QY 301 SGLTIDRQHTINWATQISWVPTIIFRSDGGATWTRI-----WD----- 342

DB 2859 SATVVPGLDGTAVA-----FRNNSPATVWGLNSSDVNTWQGSBELTQVDA 2907

QY 343 -----WTSYPNRSLRYL-----DISAEPWLTFGVQPNPVPSPKL 378

DB 2908 NSQVSLTVVDGTYVLFFTSSTASAYATSDGLNWGDITLIPW-----DDGNL 2956

QY 379 GWMD-----AMADPFNSDRMLYGTGATLYATNDLTWDSGGQIH----- 419

DB 2957 GGVASILFNQSFIILSNQSNESLLFAFNSLFPNQASRW--GEQVRDIDGDFNGDGIAD 3014

QY 420 --TAPMVKGLEEAVNDLISPPSGAPLISALGDL-----GGFTHADVAVPSTIFTSPVF 472

DB 3015 LAVLAPGRNLLQPPILDY-----PAIANNLGGVFYIYGESGISVNDPDPVVVAAPDL 3067

QY 473 TTGTSVDYAEINPSIIVRAGSFDP--SSQP-----NDRHVAFTSDGGKNW-- 515

DB 3068 PQETIPELLEITPTGVDNGDGDLLILSAPLTPVIAGQFPDVGNDQGVSWVVFSGTHGT 3127

QY 516 -FQSGEP-----GGVTTG-----GTVAASADG--SRF--VW 541

DB 3128 EYTANSPFGLNLANNQTNNSQNFYCFVTTGLPRSQAGISISGGADVNGDGFSDFALG 3187

QY 542 APGD-----PQPVVYAVGNSWAAASQ-----VPANAIQRSDR 576

DB 3188 APGNFDMLSYLVFGSDFTNQVNLGTIGDDVMLGSPTEIFVAGQGDQDIYTINGGVDTVY 3247

QY 577 VNPKTFFVSLNGFTYRSTDDG 597

DB 3248 APGNDFVTTDTNFRRLDGG 3268

# RESULT 8

S76412

hypothetical protein slr0408 - Synchocystis sp. (strain PCC 6803)

C;Species: Synchocystis sp.

A;Variety: PCC 6803

C;Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000

C;Accession: S76412

R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;

O. K.; Okumura, S.; Shimpou, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda,

DNA Res. 3, 109-136, 1996

A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synchocystis  
S.

A;Reference number: S74322; MUID:97061201; PMID:8905231

A;Accession: S76412

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-4199 <KAN>

A;Cross-references: EMBL:D90915; GB:AB001339; NID:g1653604; PIDN:BAAL8541.1; PID:g165362;

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C;Superfamily: Synchocystis hypothetical protein slr0408

Query Match 4.6%; Score 186.5; DB 2; Length 4199;

Best Local Similarity 19.2%; Pred. No. 0.019;

Matches 198; Conservative 124; Mismatches 319; Indels 389; Gaps 54;

QY 9 SNAIAGGGFGVDGIV-----FNEGAPGILYVRTDIGMYRWDAAANGRIWIPLL 55

DB 2261 ADVASDNGFVIDGNLIGNPPTTFETTSQYIDTTPAILINGSNLYLAYKFGGNNQ---IY 2317

QY 56 DWGWNW-WGVNGVVSIAAADPINTKNVW-AAVGMVYNSW-----DPNDGA-ILRSSDQA 107

DB 2318 FTVSTNGQSNSEVOL---PQSAQTIFPPALAFNNVLYAYVDGNGNLNITSQDQG 2374

QY 108 TWQITPLPFKLGNNPG-----RG-MGERLAVDPNNNDNILYFGAPSKGLWRSTDGA 159

DB 2375 TWN---APLALGGTSTPTPLFVYQGTLSLLFAANNSTVTLQFVLNSSNEIYANEIGS 2431

QY 160 TWSQMT-----NFPDVGTYIANPTDTTGYQSDIOGVVWVAF---- 195

DB 2432 NQTAISAIATVGLDGLYLVYKGGTRNTPSTLDYITSTN-----ADLSANDWSSPIPG 2486

QY 196 --DKSSSSLGQASKTIFVGVADPNNPV-FWSRDGGATWQAVCAPTGTFI---PHKGVFDP 249

DB 2487 VSSGGFSLTNDGNTNLYLSYLDSSNQLNLFVSSGNGINWSS-PQVITNNISQSPALAF-- 2543

QY 250 VNHVLYATNTSGEPYDSSGDVWKFSTGTWTRISVPVST-----DTAN 295

DB 2544 ANNELYLS-----YPGQGS-QELNVTS-----FPLPFTGSLGNSLVRFLGDVNG 2589

QY 296 DYFG--YSGLTIDRQHPNTIM-----VATQISWVPTIIFRSDGGATWTRIDWTSY 346

DB 2590 DGFADVFSGGT---NAGAIIFGNSTKDLITTAGSEDLVI-----SV 2628

QY 347 PNRSLRYL---DISAEPWLTFGV-----QPNPPVPSPKL 378

DB 2629 PNATLRDVISVGDVNGDGIKOLGVLDDGNGNFVVLGNTSLGLDKLTLSITSSSPVINVQV 2688

QY 379 GWMDEAMADPFNSDR---MLYGT-----GATLYATNDLTWMD----- 413

DB 2689 GGVTKSMAIGDYGVDVLLWGDNGNQVAGNSTGVLSFTNIDYPTQTATTGVDLN 2748

QY 414 -----SGQIHIAPMVKGLEEAVNDLISPPSGAPLISALDGLGGTHAD 458

DB 2749 SDGIPEIAGSDERKIAQISTSGSFSLPTPTTSSVINTLAAANQLENLFGNGDGIAD 2808

QY 459 VTAVPSTIFTS-----PVFTTGTSDVYAEIN-----PSIIVRAGSFDP 497

DB 2809 LAVLASNYAAIAGBPNNLPNLYSLPFGNGQGVFIYFGNSGLNSNTAQPDVILAAPPTNES 2868

QY 498 SQ-----PNDRHVAFTSDGGKNW----- 515

Db 2869 GQISTYQLSRIAQAGDVNGDFDLDLISPPYTVDAENNOGGFVVVFGDDHNNQPPFDLQ 2928  
QY 516 -----FGSEPGVTTGGTVAASA-----DGRSF-----VWAPG----- 544  
Db 2929 LRANQSGSNRPGFAIDGSPNSQAGIALNGGDIINGDFADFIICAPGENNLQYNQOIVF 2988  
QY 545 -----DPGQPVVAVG-----FGNSWAASQGVYVANAQIRSDRYN-----PKTFVAL 585  
Db 2989 IENGELSDDDKYSLYLDGNQTIOMGGGDWOANO-VWNTN-QVATNMNNSRPPPEAVIGQ 3046  
QY 586 SNGTFYRSTDCGVTFQ-----PVAAGLPSSGAVGVVMPHAFVCK 623  
Db 3047 SNGDIWYYPGQWQWQWGLKPAELINELAVNWTSGNPQIIAGLCKG--GIEYY----- 3099  
QY 624 EGDWLAASSGLYHSTNGSSW-SAITGVSSAVNVFGKSPAGSSYPVAVVGTIGGVYG 682  
Db 3100 NGSTWV--NNGPYQ---GDGWRSAITQMA---VQMGED--GS-PSQIVVLGADGAVI 3145  
QY 683 AYRSDDCGTTWVLNDDQHQYGNWQAITG-----DHANLRVYIGTN 725  
Db 3146 YYNQ---SGWRTN-----NFGKSVTLQSVQWQZASPNVIVGLDNSEV-QYYQGSN 3194  
QY 726 GRGIVYGDIG 735  
Db 3195 GVWTFPHDDG 3204

RESULT 9  
T13256  
tail-host specificity protein homolog - Lactococcus lactis phage BK5-T  
C:Species: Lactococcus lactis phage BK5-T  
C>Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 13-Aug-1999  
C:Accession: T13256  
R:Boyce, J.D.; Davidson, B.E.; Hillier, A.J.  
Appl. Environ. Microbiol. 61, 4089-4098, 1995  
A:Title: Sequence analysis of the temperate Lactococcus lactis bacteriophage BK5-T and  
A:Reference number: 217646; MUID:96064422; PMID:8526523  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1904 <BOY>  
A:Cross-references: EMBL:I44593; NID:928826; PID:928828; PIDN:AAA98579.1

Query Match 4.5%; Score 182; DB 2; Length 1904;  
Best Local Similarity 20.7%; Pred. No. 0.012;  
Matches 189; Conservative 83; Mismatches 260; Indels 380; Gaps 55;

QY 1 ATTOPYTWSNVAICGGFVDGIVNEGAPG-----ILYVRTDIGMYRWDANR 50  
Db 838 AKPSDYTWS--LIRGNDKDGATGKGVAGKDGVIKTTVITYALSSSG---TDKPTG 891  
QY 51 W---IPLLDWYGNWYGVVVSIAADPINTNKVAAVGMVTNSWDPNDGAILRSSDOGA 107  
Db 892 WTSQVPTL-----VKGQVLTWTKTWT---YIDS-----SSETG- 921  
QY 108 TWQITPLPFLKGNMGRGERLAVDPNNDNILYFGAPSGKGLWRSDGATW-SQMTN 166  
Db 922 -YSVTYI--AKDGNNGNDGIAGKGVGKTKTITYAVGTSG-----ITAPASGWNQVFN 973  
QY 167 FP-----DVGYIYANPTDTTGYQSDIQVWVWAFDKSSSLGQAKTIFVGVADPN 218  
Db 974 VPAGQFLWTKTVTYTDN-TSETGYSVAMVMGVKDGKDPGNGNTN-----GIA----- 1020  
QY 219 PVFWSRDG-----GATWQAVPG---APTG---FIP--HKGVF-----DP 249  
Db 1021 ----GKDGKGIKATITTYQASPNQTTAPTGTWSASVPPVAKGSLWTRTIWYTDNTTET 1076  
QY 250 VNHVLIATSNGTGPGYDSSGDVWKFVSTGVTWRISFPVSTDTANDYFGYSGLTIDRQH 309  
Db 1077 GYAVAYWGTNGNG-HDGFPG-----KDGTKITTTIT-----YAGTSGTTP 1118  
QY 310 PNTIMVATQISWPDPTTIFRSTDCGATWTR-IDWWTSPNRSRLAYLVDISAEPMLTFGVQ 368

Db 1119 PN-----NGWTSVTP-TVAEGNVLWTKTVWYTD--NTS----- 1149  
QY 369 PNPPVPSPKLGMDAMAI-----DPFNSDRMLYGTGATLYATNDLTKWDSCGQIHIAPM 423  
Db 1150 -----ETGYSVAMVMGVKDGKDPGNN-----GTNGIAGKDG----- 1180  
QY 424 VKGLEETAVNDLISP-PSGAPLISALDGLGGTHADVTAP--STIFTSFVFT----- 473  
Db 1181 -KGIKATAITTYQASPNQTTAPT-----GTWSASVPPVAKGSLWTRTIWYTDNTTE 1231  
QY 474 TGTSVDYAEIN-----PSIIVRAGSPDPSPQPNDRHVAFSTDGKKNWF 516  
Db 1232 TGAYAVAYWGTNGNCHDGPFGKDGKGIKTTITYAGTSGTTPPNN-----GW- 1279  
QY 517 QGSEPGVTTGGTVAASADG-----SRFVW-----APGDPGPVVVYAVG 555  
Db 1280 -----TSTVPTVAEGNVLWTKTVWYTDNTSETGYSVAMVMGVKDGKDP----- 1323  
QY 556 FGNSWAASQGVYVANAQIRSDRYNPKTFVALSNGTFYRSTDDGGVTFQVOPVAA 606  
Db 1324 -GNN--GTNGI--AGKDGKGIKATITTYQASPNQTTAPTGTWSASVPPVAKGSLWTRTIW 1379  
QY 607 -----LPSSGAVGVW-----FHAVPGKGD-----LWLAASSGLYHSTNGSS-- 644  
Db 1380 TYTDNTTETGYAVAYWGTNGNCHDGPFGKDGKGIKTTITYAGTSGTTPPNNGMTSTV 1439  
QY 645 -----W-----SAITGVSSA-VNVFGKSPAGSSYPVAVVGTIGGVGTG-- 682  
Db 1440 PTVAEGNVLWTKTVWYTDNTSETGYSVAMVMGVKDGKDPGNN-----GT-NGIAGK 1491  
QY 683 -----AYRSDCGTTWVLNDDQHQYGNWQAI-----TGDHANLRV----- 720  
Db 1492 GKGIKATITTYQASPNQTT-----APTGTWSASVPPVAKGSLWTRTIWYTDNTTE 1543  
QY 721 -----YIGTWG 726  
Db 1544 TGAYAVAYWGTNG 1555

RESULT 10  
F70846  
probable PPE protein - Mycobacterium tuberculosis (strain H37Rv)  
C:Species: Mycobacterium tuberculosis  
C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
C:Accession: F70846  
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;  
Conor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: F70846  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-2523 <COL>  
A:Cross-references: GB:AL021841; GB:AL123456; NID:g3261517; PIDN:CAA17115.1; PID:e125117;  
A:Experimental source: strain H37Rv  
C:Genetics:  
A:Gene: PPE

Query Match 4.3%; Score 175.5; DB 2; Length 2523;  
Best Local Similarity 21.8%; Pred. No. 0.044;  
Matches 175; Conservative 92; Mismatches 327; Indels 209; Gaps 38;

QY 10 NVAIGGGFVDGIVNEGA--PGILYVRTDIGMYRWDAN--GRWIPLLDWGVNWNWYN 66  
Db 1522 NTGINSGIAGTGLFNAGGFTGVVNVAGSYNTGSGNAGQANTGGFNPGSVNTGLNTG-- 1579  
QY 67 GVVSIAADPINTNKVAAVGMVTNSWDPNDGAILRSS-DOGATWQITPLPFLKGNMGP- 124  
Db 1580 -----DINTG-----VANSGDYNTGAFISGNSNGAFWR-----GDYQGL 1614

QY 125 RGMERLAVDPNDNINLYFGAPSGKGLWRSTDSGATWSQMTNFPDVGTYIANPTDTTGYQ 184  
Db 1615 LGFSYPAVLPO-----TPFLDLTLTGLGSGVVPIDPAIR 1652  
QY 185 SDIQGVVW-AEDKSSSLGQ---ASKTIIFGVAD-----PNNPVFWSRGGATWQA 232  
Db 1653 PERSANVADSFTVPISIPIDQDLAATTVSGLGPIVPHLDIPRVPVTLNYLFGSQ--- 1709  
QY 233 VPGAPTFPHKGVDPVNVHLYIATSNCTGPDGSGDVWKFVSITSGTWRISPPVSTD 292  
Db 1710 -PGPLKIGPITGLFNTPIGLTELALSOIIVIGASSQGITAFANLPFSTPVVITDEIP 1768  
QY 293 TANDYFGYS-----GLTIDRQHNTI-----MVATOISWMPDTI----- 334  
Db 1769 LLASITGHSEPVDFPGGLTIPAMNPLSINLSGCTGAVTIPAITIGBIPFDLVAHSTLG- 1827  
QY 335 ATWTRIWDTSYPNRSRLRYVLDISAEPLWTFGVQPPVPSPKLGWDE----- 383  
Db 1828 -----PVHILIDLPAVP--CFG-----NTTGAPSSGFPNFGAGVSGFNV 1866  
QY 384 -AMAIDPFN--SDRMLYGTGATLATNDLTKWDGGQIHIAPMWKGLBEETAVNDLISPPSG 441  
Db 1867 GAWVSGWQAQPSALLGGSGVFNAGTL-----HSGVLNFGSGHSGLFTSVLGL-----G 1917  
QY 442 AP-LISALGDLGFTHADVTAVPSTIPTSFTTGTSDVYAEINPSIIVRAGSDFDPSSQP 500  
Db 1918 APALVSGLSVGQ-----QLSGLLASGTA-----LHQGLVNLNGLADVGL-- 1957  
QY 501 NDRHAFSTDGKKNWFOGSEPGVTTGGTVAASADGRFV-WAP---GD-----PGQPVVY 552  
Db 1958 --GNVGLGNVDFNLGAGNVGFGNFGGNTGNNVGLNGVGNFGLGNSGLTPGLMLG 2015  
QY 553 AVFGNSMAASQGVAPANAQIRSDRVNPKTFYALSNGTF-----YRSTDDGV 598  
Db 2016 NIGFNGAGSTNFGI-ANM-----GVGNIGFANTGSGNFGI GLTGDNLTFGGFNTGSGNV 2069  
QY 599 TFOVPAAGLPSSGAVGV-MPHAVPGKEGDLWLAASSGLYHSTNGSGSSWASITGVSSAVNV 657  
Db 2070 -----GLFNSGTGNVGFNSGTGN-----WGVFNSGSY-NTGIGNSGIASTGLFNAGGF 2117  
QY 658 GFGKSAFPGSSYPAVVFVGTIGTGVAYRSDDCGTWVLINDDQHQYGNWQA-----ITG 712  
Db 2118 NTGVWAGSYNTGSPNAGQAN--TGGFNPGSVNTGWLNTGINTGVANSQDVNTGAFISG 2175  
QY 713 DHANLRVYGTNGRGIVYGDIG 735  
Db 2176 NYSN-----GAFWRGDIYQGLLG 2192

RESULT 11  
AE1905  
outer membrane secretion protein alr0791 [imported] - Nostoc sp. (strain PCC 7120)  
C:Species: Nostoc sp. PCC 7120  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
C:Accession: AE1905  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,  
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.  
DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AE1905  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-993 <KUR>  
A:Cross-references: GB:BA000019; PIDN:BA72748.1; PID:g17130136; GSPDB:GN00179  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: alr0791

Query Match 4.3%; Score 172; DB 2; Length 993;  
Best Local Similarity 19.6%; Pred. No. 0.02;  
Matches 168; Conservative 96; Mismatches 325; Indels 270; Gaps 44;

## RESULT 12

AE2254

hypothetical protein alr3588 [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp. PCC 7120

A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002

C:Accession: AE2254

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,

Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.

DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana



A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AE2254

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-908 <KUR>

A:Cross-references: GB:BA000019; PIDN:BAB75287.1; PID:g17132721; GSPDB:GN00179

A:Experimental source: strain FCC 7120

C:Genetics:

A:Gene: alr3588

Query Match 4.2%; Score 171; DB 2; Length 908;  
Best Local Similarity 20.7%; Pred. No. 0.021;  
Matches 161; Conservative 98; Mismatches 272; Indels 246; Gaps 41;

QY 71 IAADPINTNKVAAVGMVNTSWDPNDGAILRSSDQG---ATWQITPLPFKLGNNMPGRGM 127

DB 131 LSAEVIQNNI-----NDNTIANARNIGTNTTITPVDYVSGS----- 171

QY 128 GERLAVDPNDNIIYFGAPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDTTGYQSDI 187

DB 172 GQ--VIDQND---YIS-----FTLTNSGTV-----DIINLSLNGTDT--LYADL 208

QY 188 QGVVVAFAFDKSSISLGOASKTIFVGVAADPNPVFWSRDGGATQAVP---CAPTGFIPIHK 244

DB 209 QLI-----SSSNVIQTSATV-----GTSLESISLSLAAGTYIIRAY 245

QY 245 GVDFPVNHLXI-----ATSNLTGGPYD-----GSSGDVWKFSVT 278

DB 246 SOSDPGNVLEFNFSADPPDAGGNTSDTSTINLPATFSEIISDOVSLGSDSYQFTLA 305

QY 279 SGWTRISPVSTDTANDYFGYGLTIDRQHPNTIMVATQISWPDITIIIRSTDG----- 333

DB 306 SASLVEIQFTSLTADANLY-----LQTONGGINLSTQPGTALDAVRLSLNAGTYNII 358

QY 334 --CATWRIWDWISY-----PNRSL-----RYVLDLSAEPLWTFGV 367

DB 359 VNRGSTETAQITLGSFAQIAGNDQAPNSTTIALNLINSPISLNEFVGNIDTNDYIKPTV 418

QY 368 QPNPVPSPKLGMD---EAMADP---FNSDRMLYGTGATLYATNDLTWKDS----- 414

DB 419 NGTTEIN-----LDLSILNSYLLDPQLVADVQILNSGGTQVAISNQTGNSNESINTIL 472

QY 415 -GGQIHAPWKVLEETAVNDLISPPSGAPILSALGDLGGFTHADVTAVPSTIFTSPVFT 473

DB 473 GAGTYFIRVYTSGLANTFYDLNITAQSOALLVQDINPTGNSDPANLTLTGLNTLY---FT 528

QY 474 TGTSDVDAEL---NPSIIIV---AGSPDPSSQ---PNDRHVAFSTD--GGKMW-FQG--- 518

DB 529 ANDGINGVOLWSSNGDITRLSNISSEFNPTNLIVFNNRLYFAASNDTFGRELWEYNGTV 588

QY 519 -----SRPGGVTTGGT---VAASADGSRFWA-PGDPGQPVVYAVGFGNSWA 561

DB 589 NRISDINVAGNSNPGLTVAGNKLFFTAVDNDSIRKLWYNGTNVNLVDVNASFSNSST 648

QY 562 ASQGVPAQAQIRSRVNPCKTFFYALNSGTYFYRSTDG--GVTFQPVAAAGLPSSGAVGVMPHA 619

DB 649 PTFETTFNNQL-----FFTAQNSQLWSTDGTIGTQVISAGGITKSTPRNL---T 696

QY 620 VPGKEGDLWLAAASGLYHSTNGGSSWSAITGVSSA-----VNVGFKSAPGSSYPAVFVG 675

DB 697 VVG--NTLYFTANRG---TSGHEIWQYQNGTTASLLEDITFGNNSFAP--SSLTAV----- 745

QY 676 TIGGVTGAYRSDDCGTTWVLNDQHOYG--NW---GOAITGDHANLRVRVIGNIGR 727

DB 746 -----GNTLYFVTDSDNDFNLWLKWSGDTGAAGTD-----IIGTDGQ 781

#### RESULT 13

B85547

probable RTX family exoprotein [imported] - Escherichia coli (strain O157:H7, substrain

C:Species: Escherichia coli

C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001

C:Accession: B85547

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca, Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: B85547  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-5188 <STO>  
A:Cross-references: GB:AE005174; NID:g12513369; PIDN:AAG54838.1; GSPDB:GN00145; UWGP:Z061  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: Z0615

Query Match 4.2%; Score 170.5; DB 2; Length 5188;  
Best Local Similarity 20.3%; Pred. No. 0.22;  
Matches 191; Conservative 82; Mismatches 269; Indels 401; Gaps 47;

QY 27 GAPGILYVRTDIGMYRWDAANGRWIPLLDWGVNNWNGVGVVSIADAPINNKVAAVVG 86

DB 4083 GANEPAQISTDNGA-----TWVN-----VTVAADSLN----- 4109

QY 87 MYTNSWDNDGAILRSSDQATWQITPLPFKLGNNMPGRGMLERLAVDPNN----- 137

DB 4110 -----WSYVDGRTLNGT--TWQVRVV--DIAGNV-GATSSQSALIDTVNPAQVLTIAS 4159

QY 138 -----DNILYF-----GAPSGKGLWRSTDGATWSQMTNFPDVGTYIAN 176

DB 4160 ISTDGSSATDFTISDTWLTGSLGAGLASGEVAQISLDSGATWTLTTNGTQWYITDS 4219

QY 177 PTDTTG---YQSDIQGVVWVAFDKSSSLGOASKTIFGVADPN---NPVFWSRDGGATW 230

DB 4220 RLTLDGSVYQVRV-----LDLAGNTGPVVKTVVVDTPINPTATPTIVSYTDDVG--- 4269

QY 231 QAVPCAPTGFPHKGVFDPVNVHLYIATSNCGPYDG-----SSGDV----- 272

DB 4270 -----ORQITSLSLQ-----ATDDTTPLLNGVLASPLASGEVYLYRNGLLGA 4313

QY 273 -----WKFS---VTSGTWT---RISFPVSTDTANDYFGYSGLTIDRQHPNTIMVAT- 317

DB 4314 VTMVGCALNWTYSDSLVSGAYTYSARVVDLAGNITSSDF---VLTVDTSIPTTLAQITS 4370

QY 318 -----QTSW---PPT--- 325

DB 4371 QTRDTPPIISGVITAALASQYVEWINGKTYTSEPGGVAVVVDPAHNTWYQLPDTDAL 4430

QY 326 -----IIFRSTDGATWTRWDWTSYPNRSIRYVL 355

DB 4431 TVSATAYTVAQVKSAGNANNANISNGTIVVNAADYTPPTWTTASKTTAW---GLTYGL 4487

QY 356 DISAEPWLTFGVQPNPPVPSPKLGWMDAMADPPNSDRMLYGTGATLYATNDLTWKDSG 415

DB 4488 D-SHGWTVLANQQVMQSTDP-LTWSKTALT-----LYQSGNN-YATSIADYDRN 4535

QY 416 GQIHIAPMVKGLEETAVNDLISPPSGAPLIS---ALGDLGGFT-HADVTAVPSTIFTSP 470

DB 4536 G-----TGDLFITRDDYGTGYINGFTNNGDGT-----FSSA 4566

QY 471 V-FTTGTSVDYAEALNPSIIIVRAGSFPDPSQPNDRHVAFST---DGGKNWTFQSGEPG- 522

DB 4567 IQVTVGLTLWTGSI-----VAFDKEGDGYLDFEWIGDAGGPDST 4605

QY 523 -----GVTTGCTVAASADGSRFWAPGDPGQPVVYAVGFGNSWAASQCVPAQAQIRSR 576

DB 4606 FLNNAGTLVNGSTTSNGGSATVGG-----AVTGYLSLNEGSGVDLNNDRIDL 4655

QY 577 V---NPKTFYALS-----NGTFYRSTDGGVTQFQFAAGLPSSGAVGVYMFHAVPGK-BG 625

DB 4656 VQHTVNLNNYITLSSLINQNGTFFWQGNNTNTFLSGAGSGAWSSVSWTWADFDGDM 4715

QY 626 DLWLAAASSG-----LYHSTNG----- 641

DB 4716 DLFLPASQGRANYGSLLEFNTNGVLGCPVAVGATATTYASQFSLAVDWNHDLGLDIARIAQ 4775

Qy	471	V-FTTGSVDYAEINPISIIVRAGSFDPSQPNDRHVAFT--DGGKNWFGCSSEPG-----	522
Db	4670	IQVTVGTLTWYGSF-----VAFDKREGDGYLDFTWIGDAGGPDNST	4708
Qy	523	-----GVTTGGTVAASADGSRFWAPGDPQPVWVAVGFGNSWAASQGVPAQAIRSDR	576
Db	4709	FLNNAGTLVGNSTTSNGGSATVGG-----AVTGYLSINEGSGVDLNDGRIDL	4758
Qy	577	V-----NPKTFYALS-----NGTFYRSTDGGVTFPQVAAGLPSSGAVGVNMFHAVPGK-EG	625
Db	4759	VQHTYLNYYTLLSSLLNQGNGTFVWGQNTTNTFLSGAGSGAMSSSVSMTWADFDGDCDM	4818
Qy	626	DLWLAASSG-----LYHSTNG-----	641
Db	4819	DLFLPASQGRANYGSLLENTFNGVLGCPVAVGATVTVASQFSLAVDNHNDGLMDIARIAQ	4878
Qy	642	GSWSAITGVSSAVNVGFKSPAGSSYPAVFVVVGTTGGVT-----GAYR---SDDCG	690
Db	4879	TGQSY-LYTVNVSASN--WTQSALGSGQS-----GTTSGVAAMDYDWDGAVDLVSKQSG	4930
Qy	691	TTWVYLINDHQHVGWNGQAITGDHANLRVYVYGTNGRGIVYGD	733
Db	4931	SVFLSRNTNTVSYG-----TSLHLRI-----TDPNGINVIYGN	4963
RESULT 15			
A28452			
Proteoglycan core protein precursor, cartilage - rat			
N:Alternate names: aggrecan			
C:Species: Rattus norvegicus (Norway rat)			
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 21-Jan-2000			
C:Accession: A92623; A28335; A28453; A28095; A28452			
J:Doerge, K.; Sasaki, M.; Horigan, E.; Hassell, J.R.; Yamada, Y.			
J: Biol. Chem. 262, 17757-17767, 1987			
A:Title: Complete primary structure of the rat cartilage proteoglycan core prot			
A:Reference number: A92623; MUID:88087070; PMID:36933370			
A:Accession: A92623			
A:Molecule type: mRNA			
A:Residues: 1-2124 <DOE>			
R:Doerge, K.; Sasaki, M.; Horigan, E.; Hassell, J.R.; Yamada, Y.			
J: Biol. Chem. 263, 10040a, 1988			
A:Reference number: A30069			
A:Contents: annotation; revision to residue 698			
R:Doerge, K.; Fernandez, P.; Hassell, J.R.; Sasaki, M.; Yamada, Y.			
J: Biol. Chem. 261, 8108-8111, 1986			
A:Title: Partial CDNA sequence encoding a globular domain at the C terminus of			
A:Reference number: A23835; MUID:86250698; PMID:2424893			
A:Accession: A23835			
A:Molecule type: mRNA			
A:Residues: 1856-2124 <DO2>			
A:Cross-references: GB:M13518; NID:g206104; PIDN:AAA41836.1; PID:g206105			
R:Neame, P.J.; Christner, J.E.; Baker, J.R.			
J: Biol. Chem. 262, 17768-17778, 1987			
A:Title: Cartilage proteoglycan aggregates. The link protein and proteoglycan a			
A:Reference number: A28453; MUID:88087071; PMID:36933371			
A:Accession: A28453			
A:Molecule type: protein			
A:Residues: 20-37, 'W', 39-60, 'E', 62-64, 'X', 66-69, 70-83, 84, 89-148, 'L', 150-238, 'S',			
C:Superfamily: aggrecan; C-type lectin homology; complement factor H repeat hom			
C:Keywords: Glycoprotein			
F:1-19/Domain: signal sequence #status predicted <SIG>			
F:20-2124/Product: proteoglycan core protein #status predicted <MAT>			
F:44-135/Domain: immunoglobulin homology <IMM>			
F:170-247/Domain: link protein repeat homology <LNK1>			
F:268-349/Domain: link protein repeat homology <LNK2>			
F:504-581/Domain: link protein repeat homology <LNK3>			
F:602-683/Domain: link protein repeat homology <LNK4>			
F:1914-2034/Domain: C-type lectin homology <LCH>			
F:2941-2097/Domain: complement factor H repeat homology <FHD>			
F:126, 239, 333, 387, 611, 667, 1842/Binding site: carbohydrate (Asn) (covalent) #sta			
Query Match			
4.2% Score 168; DB 2; Length 2124;			



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 11:41:35 ; Search time 20.1745 Seconds  
(without alignments)

1909.933 Million cell updates/sec

Title: US-09-917-376-3

Perfect score: 4036

Sequence: 1 ATTQPYTWSNVAIGGGGFVD.....YIGTNGRGIYGDIGGAPSG 740

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	168	4.2	2124	1 PGCA RAT	P07897 rattus norv
2	167.5	4.2	1119	1 ALS3 CANAL	O74623 candida alb
3	166	4.1	1468	1 N153 RAT	P49791 rattus norv
4	157	3.9	809	1 QUIA ACICA	Q59086 acinetobact
5	154.5	3.8	5703	1 MUSB HUMAN	Q3hc84 homo sapien
6	153	3.8	1150	1 APMU PIG	P12021 sus scrofa
7	152	3.8	1196	1 AMYE BAEPO	P21543 paenibacill
8	149	3.7	2132	1 ISOA PSBAY	Q61282 mus musculu
9	148.5	3.7	776	1 ISOA PSBAY	P10342 pseudomonas
10	148.5	3.7	776	1 ISOA PSBAY	P26501 pseudomonas
11	148.5	3.7	1953	1 BIGA SALTY	P25927 salmonella
12	146	3.6	2109	1 PGCA CHICK	P07898 gallus gall
13	145.5	3.6	790	1 QUIA XANCJ	Q9xd78 xanthomonas
14	145	3.6	342	1 Y48L SYNV3	P73069 synechocyst
15	145	3.6	747	1 GUND CELFI	P50400 cellulomona
16	144.5	3.6	708	1 QHED COMTE	Q46444 comamonas t
17	143.5	3.6	937	1 NU98 RAT	P49793 rattus norv
18	142.5	3.5	995	1 AGAA VIBES7	P48839 vibrio sp.
19	142	3.5	647	1 NANH MICVI	Q02834 micromonosp
20	142	3.5	757	1 DBET GLUOX	O05542 gluconobact
21	142	3.5	760	1 YH1L ECOLI	P75780 escherichia
22	139.5	3.5	1045	1 GUNB CELFI	P26225 cellulomona
23	139.5	3.5	2333	1 PGCA CANFA	Q28343 canis famil
24	139	3.4	3063	1 CALC HUMAN	Q99715 homo sapien
25	138.5	3.4	955	1 YC6G ECOLI	P76017 escherichia
26	138.5	3.4	1526	1 Y4V6 ANASP	Q8yri1 anabaena sp
27	138	3.4	827	1 XANP XANS2	Q60106 xanthomonas
28	138	3.4	1159	1 SOR2 MOUSE	Q9ep35 mus musculu
29	137.5	3.4	872	1 GUXA CELFI	P50401 cellulomona
30	137.5	3.4	1103	1 VG37 BPARI	Q9g0b5 bacterioph
31	137.5	3.4	2415	1 PGCA HUMAN	P16112 homo sapien
32	136.5	3.4	522	1 NU62 HUMAN	P37198 homo sapien
33	136.5	3.4	781	1 NANH_VIBCH	P37060 vibrio chol

34	136.5	3.4	1729	1 NU98 HUMAN	P52948 homo sapien
35	136	3.4	2628	1 HGA2_PORGI	Q51845 porphyromon
36	135.5	3.4	725	1 YA33 SULSO	Q97297 sulfolobus
37	134.5	3.3	2344	1 POLN_RHDV	P27410 rabbit hemo
38	134	3.3	890	1 BCN5_CLOPE	P08696 clostridium
39	134	3.3	1014	1 NANH_CLOS	P29767 clostridium
40	133.5	3.3	966	1 FIB1_PETWA	P02674 petromyzon
41	132	3.3	886	1 VGP3_EBVAB	Q07284 epstein-bar
42	132	3.3	1433	1 SUBF_BACSU	P16397 bacillus su
43	132	3.3	1541	1 ASX1 HUMAN	Q81xj9 homo sapien
44	131.5	3.3	1026	1 VG37_BPT4	P03744 bacterioph
45	131.5	3.3	1199	1 P121_RAT	P52591 rattus norv

#### ALIGNMENTS

RESULT 1  
PGCA RAT  
ID PGCA RAT STANDARD; PRT; 2124 AA.  
AC P07897;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Aggrecan core protein precursor (Cartilage-specific proteoglycan core protein) (CSPCP).  
GN AGC1 OR AGC.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
NCBI\_TaxID=10116;  
RX MEDLINE=8087070; PubMed=3693370;  
RA Doege K., Sasaki M., Horigan E., Hassell J.R., Yamada Y.;  
RT "Complete primary structure of the rat cartilage proteoglycan core protein deduced from cDNA clones."  
RL J. Biol. Chem. 262:17757-17767(1987).  
RN [2]  
RP REVISION TO 698.  
RA Doege K., Sasaki M., Horigan E., Hassell J.R., Yamada Y.;  
RL J. Biol. Chem. 263:10040-10040(1988).  
RN [3]  
RX MEDLINE=86250698; PubMed=242493;  
RA Doege K., Fernandez P., Hassell J.R., Sasaki M., Yamada Y.;  
RT "Partial cDNA sequence encoding a globular domain at the C terminus of the rat cartilage proteoglycan."  
RL J. Biol. Chem. 261:8108-8111(1986).  
CC -!- FUNCTION: This proteoglycan is a major component of extracellular matrix of cartilaginous tissues. A major function of this protein is to resist compression in cartilage. It binds avidly to hyaluronic acid via an amino-terminal globular region. May play a regulatory role in the matrix assembly of the cartilage.  
CC -!- SUBUNIT: Interacts with FBLN1 (By similarity).  
CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix (By similarity).  
CC -!- DOMAIN: Two globular domains, G1 and G2, comprise the amino terminus of the proteoglycan, while another globular region, G3, makes up the COOH terminus. G1 contains link domains and thus consists of three disulfide-bonded loop structures designated as the A, B, B' motifs. G2 is similar to G1. The keratan sulfate (KS) and the chondroitin sulfate (CS) attachment domains lie between G2 and G3.  
CC -!- PTM: Contains mostly chondroitin sulfate, but also keratan sulfate chains, N-linked and O-linked oligosaccharides.  
CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.  
CC -!- SIMILARITY: Contains 4 link domains.  
CC -!- SIMILARITY: Contains 1 C-type lectin family domain.  
CC -!- SIMILARITY: Contains 1 Sushi (SCR) domain.  
CC -!- SIMILARITY: Belongs to the aggrecan/versican proteoglycan family.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration



RESULT 2  
 ALS3 CANAL  
 ID ALS3 CANAL STANDARD; PRT; 1119 AA.  
 AC 074623;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Agglutinin-like protein 3 precursor.  
 GN ALS3.  
 OS Candida albicans (Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
 OX NCBI\_TaxID=5476;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1161.  
 RX MEDLINE=98309840; PubMed=9644209;  
 RA Hoyer L.L., Payne T.L., Bell M., Myers A.M., Scherer S.;  
 RT "Candida albicans ALS3 and insights into the nature of the ALS gene  
 family.";  
 RL Curr. Genet. 33:451-459(1998).  
 CC -!- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.  
 CC -!- PFM: N-glycosylated and O-glycosylated (Potential).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U87956; AAC39486.1; --  
 DR InterPro; IPR008440; Candida ALS.  
 DR Pfam; PF05792; Candida ALS; 1.  
 KW Cell adhesion; Glycoprotein; Repeat; Signal.  
 FT SIGNAL 1 17  
 FT CHAIN 18 1119  
 FT DOMAIN 433 792  
 FT REPEAT 433 468  
 FT REPEAT 469 504  
 FT REPEAT 505 540  
 FT REPEAT 541 576  
 FT REPEAT 577 612  
 FT REPEAT 613 648  
 FT REPEAT 649 684  
 FT REPEAT 685 720  
 FT REPEAT 721 756  
 FT REPEAT 757 792  
 FT DOMAIN 399 404  
 FT DOMAIN 450 455  
 FT DOMAIN 557 563  
 FT DOMAIN 593 597  
 FT DOMAIN 630 635  
 FT DOMAIN 666 671  
 FT DOMAIN 702 707  
 FT DOMAIN 738 743  
 FT DOMAIN 774 777  
 FT DOMAIN 1044 1047  
 FT CARBOHYD 471 471  
 FT CARBOHYD 543 543  
 FT CARBOHYD 579 579  
 FT CARBOHYD 651 651  
 FT CARBOHYD 687 687  
 FT CARBOHYD 723 723  
 FT CARBOHYD 759 759  
 FT CARBOHYD 845 845  
 FT CARBOHYD 987 987  
 FT CARBOHYD 1050 1050  
 FT CARBOHYD 1061 1061  
 FT SEQUENCE 1119 AA; 119927 MW; 6A3FB3FC8C879A71 CRC64;

Query Match 4.2%; Score 167.5; DB 1; Length 1119;  
 Best Local Similarity 18.7%; Pred. No. 0.03;  
 Matches 170; Conservative 98; Mismatches 340; Indels 303; Gaps 42;  
 QY 7 TMSNVAIGGGFVGIVNEGAPGLYVTRTDIGG-----MYRWDANGRW--IPL-LDWW 58  
 Db 264 TYKNVPAGYRPFVDAYI-----SATDVNSYTLSEYATCAGGYQWRAPFTLRWT 313  
 QY 59 GWNN--WGYNGVSVIAADPINTNKVAAVGMTNSWDNDGAILRSDQCATWQI--TPLP 115  
 Db 314 GYRNSDAGSNGIVIVATRTVDSITAVT--TLPPDPN-----RDKTKTIELKPIP 363  
 QY 116 -----FKLGNNMPGRGMRERLAVDPNDNLIYFCAPSGKGLWRSTDSAT----- 160  
 Db 364 TTTITTSYGVVTSYSTKTAPIGETATVIV--DIPYHTTTVTTSKWTGTTTTHTNP 420  
 QY 161 -----WSQMTNFPDVGIVIANP--TD-----TTG 182  
 Db 421 TDSIDTVIVQVSPNPNTVTTEYWSQ--SFATTTITGPPGNTDTVLIREPPNHTVTTE 478  
 QY 183 YOSDIQGVVWVAFDKSSSSSLGQASKTIFGVADPNPV-----FMSRD--GGATWQAVPG 235  
 Db 479 YWSE-----SYTTSTFTAPPGTDSVIKEPPNPVTVTTEYWSESYTTSTFTAPPG 531  
 QY 236 APTGFIPIHKGVPDPVNHVL-----YIATSNLTGGPYDG-----SSGDV 272  
 Db 532 GTDSVI---IKEPPNHTVTTEYWSQSYTTTITVAPPGGTDTVLREPPNHTVTTEY 587  
 QY 273 WKFSVTSWTTRISVPSTP-----TANDYFYSGLT-----ID 306  
 Db 588 WSQSYTT--TTTIVAPPGGTDSVIREPPNPVTVTTEYWSQSYATTTITAPPGETDTVL 646  
 QY 307 ROHPNTINVAQISNM-----PDTIIRSTDGATWTRIDWTSYENRS 350  
 Db 647 REPPNHTVTTE--YWSQSYATTTITAPPGETDTVLIREPPNHTVT-----TEWSQS 699  
 QY 351 LRYVLIDISAEPLWTFGV---QPNPPVSPKLGWMD-----AMADPNSDRML----- 396  
 Db 700 FATTTTVPAPGGTDTVLIREPPNHTVTTEY--WSQSYATTTITAPPGETDTVLIREPP 758  
 QY 397 -----YGTGATLVATNDLT-----WDSGQIHIAPMVKG 426  
 Db 759 NHTVTTEYWSQSYATTTITAPPGETDTVLIREPPNPVTVTTEYWSQ-----806  
 QY 427 LEETAUNDLISPPSCAPLI-----SALDGLGFT-----HADTVAPSTIFTSPVTTG 475  
 Db 807 -SYTTATTVPAPGGTDTVIYDWTSSSEISSFSRPHYTNHTTLMSTTWIETKITETS 865  
 QY 476 TSVD---YAEINPSIIVRAGSFPSSQPNDRHAFSTDGKNWFQGSBPGGVTTGTVA 531  
 Db 866 CEGDKGCSWVSSTRIVIPNNIETPMVTN-----TVDSSTTESQSPSGIFSESGVS 919  
 QY 532 ASADGSRFVWAPGDPGPQ-----VYVAVFGNSMAASQGVANQAQIRSDRNPKTFYALS 586  
 Db 920 VETESSTVTTAQTNPSPVPTSEVVF-----TTKNNENGPVESPSTNVKSSMD-E 969  
 QY 587 NGTFIRSTDGGVTFQFVAAAGLPSGCAVGMFVAPCKEDLMLAASSGLYHSTNGSSWS 646  
 Db 970 NSEFTTSTAATSTDIENETIATGSGV-----EASSPFISSADETTTV 1013  
 QY 647 AITGVSSAV-----NVGFGKSPAGSSYPVAVFVVGITGGVTCAYRSDDCQTTWVLNDDQ 700  
 Db 1014 TTTAESTSVIEQPTNNGGGRAPSAITSPSTTTTANDSVI-----TGTSTNSQSQ 1066  
 QY 701 HQYGNWQQAIT 711  
 Db 1067 SQYNSDTQOTT 1077  
 RESULT 3  
 N153 RAT  
 ID N153\_RAT STANDARD; PRT; 1468 AA.



CC -|- COFACTOR: PQQ.  
 CC -|- PATHWAY: Quinic acid catabolic pathway; first step. This pathway  
 CC allows growth of bacteria with quinate by its conversion to  
 CC protocatechuate and subsequent metabolism by the beta-ketoadipate  
 CC pathway.  
 CC -|- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -|- INDUCTION: By protocatechuate.  
 CC -|- SIMILARITY: Belongs to the bacterial PQQ dehydrogenase family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcements/>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC  
 CC EMBL; L05770; AAC37161.1; -;  
 CC PIR; A55547; A55547.  
 CC InterPro; IPR001479; Bac\_PQQ.  
 CC InterPro; IPR002372; Bac\_PQQ\_repeat.  
 CC Pfam; PF01011; PQQ; 4.  
 CC SMART; SM00564; PQQ; 3.  
 CC PROSITE; PS00363; BACTERIAL\_PQQ\_1; FALSE NEG.  
 CC PROSITE; PS00364; BACTERIAL\_PQQ\_2; FALSE NEG.  
 CC Oxidoreductase; PQQ; Quinate metabolism; Transmembrane.  
 CC TRANSMEM 14 34 POTENTIAL.  
 CC TRANSMEM 41 61 POTENTIAL.  
 CC TRANSMEM 68 88 POTENTIAL.  
 CC TRANSMEM 90 110 POTENTIAL.  
 CC TRANSMEM 127 147 POTENTIAL.  
 CC TRANSMEM 809 AA; 88196 MW; 7167CEBEA62BFCB CRC64;  
 CC SEQUENCE  
 CC  
 CC Query Match 3.9%; Score 157; DB 1; Length 809;  
 CC Best Local Similarity 20.5%; Pred. No. 0.081;  
 CC Matches 168; Conservative 90; Mismatches 254; Indels 306; Gaps 45;  
 CC  
 CC QY 30 GILYVRTDIGGMY-----RWDAANGRIWPLL-----DWGVNNGYN--GVVSTAADP 75  
 CC DB 131 GGLTVUGMLGGLGCMPIPHETVKAAGEELPLVPDPAKKVNDWDHNDAGGSRFVALDQ 190  
 CC QY 76 INTNKVAAVGMVYTNWDPNPDNGAILRSSDQAGTQWITPLPFLKGGNPGRMGERLAVDP 135  
 CC DB 191 INENN-----VSKLEAWRFRTGDTTCTGNGAEDQMTPLQ-----VGNKVFLLCT 235  
 CC QY 136 NNDNLYFGAPSGKGLWRS-----TDSGAT----- 160  
 CC DB 236 PHNNIFAIDADSQGLWKAENVNSTADAWERCGRGVAVFDSQPLVQPTLAGATPVAALAN 295  
 CC QY 161 -----WSQMTNFPDVGTYIANPTDTTGTQSD--IQGVVWVAFDKSSSLGQASK----- 207  
 CC DB 296 TECPRRVYNTVD--GRLLIADVNTGACRCKDFGVNGTV-----NLHEGLGENTKAPREVT 349  
 CC QY 208 -----TIFVG-----VADPNPVPFWSRD---GGATWQAVPGAPTGFPHKGVFDP 249  
 CC DB 350 SAPTIAGTIIVGSRADNVAAADMPGVIKAYDVITGKLKAWDPNP-----DP 399  
 CC QY 250 VNHVLIATSNTPGPGYDGGSDVWKFVSTGVTWTRISPVSTDT-----ANDYFGYS 301  
 CC DB 400 -NVVL-----KPGEIVKRSSTN--SWAAMSVDPMQNTVFLPMGSSSVVWVGGN 444  
 CC QY 302 GLTIHQHPNTINVAQISWPDPTIIFRSTDDGATW-----TRINDWTSYPNRSRLRYVL 355  
 CC DB 445 RTAADHKY-NTSLALD-----ATTGKEKVVYNTVHNDLWDF----- 480  
 CC QY 356 DISAEPWLT-FGVQPNPPVPSKLGWDEAMAIDPFNSDRMLYGTGATLVATNDLTWDS 414  
 CC DB 481 DLPNQSLVDVFPKWDGTTKFAVIG-----TK----- 507  
 CC QY 415 GQGIHI-----APMVKGLSE-----TAVNDLISPPSGAPLISALGDLGGF 454  
 CC DB 508 SGQFYVLDRVTGKPLTKVIEQPIKVAIDPGEQVSKTQPRSVEMPIQGNQTLKE-SDMWGA 566

QY 455 THADVTAVP-----STIETSPVFTTGTSDVYA-----ELNPSIIIVRAGSEDPSSQ- 499  
 DB 567 TFPDQLMCRINFKSMRYDGLYAP-----GTDVLSLSPGSLGGMWGS-----AFDPTHY 618  
 QY 500 -----PNDRHVAFTDGGKWNWFGSGEPGGVTTGCTVAASADGSRFVW 541  
 DB 619 MFVNDMLRLGLWLIQKTPED--IKIQANGGEKVNTGM--GAVPMKGT-PYKVKRFRFMS 673  
 QY 542 APGDPGQGVVYVAVGFGNSWAASQGVPAANAQIRSDRVNPKTFYALNSGNTFYRSTDDGVTF- 600  
 DB 674 ALGIP-----CQPPFGTMTAIDMKTRQVAVQVPLGTIQTDTGPMGIKMG 717  
 QY 601 --QPVAGLPSGAV-----GVMFHAVPKGGLDLAASSGLYHSTNGSSSVAITGV-S 652  
 DB 718 LKAPI--GMPTIGPMATQGGGLVFPA--TQDYILRA-----FNSSNGKELMKARLPVGS 768  
 QY 653 SAVNVGCKSAPGSSYPVAVFVVGTTGGVTGAYRSDDCG 690  
 DB 769 QGTPMSYMSPKTGKQY-----VVVSAG---GARQSPDHG 799  
 CC  
 CC RESULT 5  
 CC MUSB\_HUMAN  
 CC ID MUSB\_HUMAN STANDARD; PRT; 5703 AA.  
 CC AC Q9HC84; O00447; O00573; O14985; O15494; O95291; O95451; Q14881;  
 CC AC Q99552; Q9UE28;  
 CC DT 28-FEB-2003 (Rel. 41, Created)  
 CC DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 CC DE Mucin 5B precursor (Mucin 5 subtype B, tracheobronchial) (High  
 CC DE molecular weight salivary mucin MGI) (Sublingual gland mucin).  
 CC GN MUC5B OR MUC5.  
 CC OS Homo sapiens (Human).  
 CC OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 CC OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 CC OX NCBI\_TaxID=9606;  
 CC RN [1]  
 CC RP SEQUENCE OF 1-1594 FROM N.A.  
 CC RA Chen Y., Di Y.P., Wu R.;  
 CC RT "Molecular cloning of the amino-terminal and 5'-flanking region of the  
 CC RT human MUC5B mucin gene.";  
 CC RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
 CC RN [2]  
 CC RP SEQUENCE OF 1-1325 FROM N.A.  
 CC RX MEDLINE=98009274; PubMed=9790959;  
 CC RA Offner G.D., Nunes D.P., Keates A.C., Afchal N.H., Troxler R.F.;  
 CC RT "The amino-terminal sequence of MUC5B contains conserved  
 CC RT multifunctional D domains: implications for tissue-specific mucin  
 CC RT functions.";  
 CC RL Biochem. Biophys. Res. Commun. 251:350-355(1998).  
 CC RN [3]  
 CC RP SEQUENCE OF 40-1324 FROM N.A.  
 CC RX MEDLINE=98023932; PubMed=9804771;  
 CC RA Desseyn J.-L., Buisne M.P., Porchet N., Aubert J.-P., Laine A.;  
 CC RT "Genomic organization of the human mucin gene MUC5B: cDNA and genomic  
 CC RT sequences upstream of the large central exon.";  
 CC RL J. Biol. Chem. 273:30157-30164(1998).  
 CC RN [4]  
 CC RP SEQUENCE OF 1326-4895 FROM N.A.  
 CC RC TISSUE=Placenta;  
 CC RX MEDLINE=97166151; PubMed=9013550;  
 CC RA Desseyn J.-L., Guyonnet-Duperat V., Porchet N., Aubert J.-P.,  
 CC RA Laine A.;  
 CC RT "Human mucin gene MUC5B, the 10.7 kb large central exon encodes  
 CC RT various alternate subdomains resulting in a super-repeat. Structural  
 CC RT evidence for a lip15.5 gene family.";  
 CC RL J. Biol. Chem. 272:3168-3178(1997).  
 CC RN [5]  
 CC RP SEQUENCE OF 4057-4480 FROM N.A.  
 CC RC TISSUE=Salivary gland;  
 CC RX MEDLINE=97292540; PubMed=9147051;  
 CC RA Nielsen P.A., Bennett E.P., Wandall H.H., Therkildsen M.H.,  
 CC RA Hannibal J., Clausen H.;



RT "Identification of a major human high molecular weight salivary mucin (MGI) as tracheobronchial mucin MUC5B.";  
 RL Glycobiology 7:413-419(1997).  
 RN [6]  
 RP SEQUENCE OF 4721-5703 FROM N.A.  
 RC TISSUE=Gall bladder;  
 RX MEDLINE=97293229; PubMed=9164870;  
 RA Keates A.C., Nunes D.P., Aidhal N.H., Troxler R.F., Offner G.D.;  
 RT "Molecular cloning of a major human gall bladder mucin: complete C-terminal sequence and genomic organization of MUC5B.";  
 RL Biochem. J. 324:295-303(1997).  
 RN [7]  
 RP SEQUENCE OF 4809-5687 FROM N.A.  
 RC TISSUE=Sublingual gland;  
 RX MEDLINE=96125355; PubMed=8554565;  
 RA Troxler R.F., Offner G.D., Zhang F., Iontcheva I., Oppenheim F.G.;  
 RT "Molecular cloning of a novel high molecular weight mucin (MGI) from human sublingual gland.";  
 RL Biochem. Biophys. Res. Commun. 217:1112-1119(1995).  
 RN [8]  
 RP SEQUENCE OF 4859-5703 FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=97347489; PubMed=9201995;  
 RA Desseyn J.-L., Aubert J.-P., Forchet N., Laine A.;  
 RT "Genomic organization of the 3 region of the human MUC5B mucin.";  
 RL J. Biol. Chem. 272:16873-16883(1997).  
 CC -!- FUNCTION: Salivary mucin that is thought to contribute to the lubricating and viscoelastic properties of whole saliva.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- TISSUE SPECIFICITY: Expressed mainly in bronchus glands and also in submaxillary glands, endocervix, gall bladder, and pancreas.  
 CC -!- PTM: Highly glycosylated.  
 CC -!- SIMILARITY: Contains 1 TIL (Trypsin inhibitory-like) domain.  
 CC -!- SIMILARITY: Contains 3 WFC domains.  
 CC -!- SIMILARITY: Contains 4 WFD domains.  
 CC -!- SIMILARITY: Contains 1 C-terminal cysteine knot-like (CTCK) domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/ebis-sib.ch/announcements> or send an email to [license@ebi-sib.ch](mailto:license@ebi-sib.ch)).  
 CC -----  
 DR EMBL: AF107890; AAC33673.1; -;  
 DR EMBL: AF086604; AAC87545.1; -;  
 DR EMBL: AJ004862; CAA06167.1; -;  
 DR EMBL: Z72496; CAA96577.1; -;  
 DR EMBL: X74955; CAA52910.1; -;  
 DR EMBL: X63836; AAB61398.1; -;  
 DR EMBL: U78554; AAC51344.1; -;  
 DR EMBL: U78552; AAC51344.1; JOINED.  
 DR EMBL: U78553; AAC51344.1; JOINED.  
 DR EMBL: U78551; AAC51343.1; -;  
 DR EMBL: U95031; AAB65151.1; -;  
 DR EMBL: Y09788; CAA70926.1; -;  
 DR EMBL: HGNC:7516; MUC5B.  
 DR MIM: 600770; -;  
 DR GO: GO:0005515; F:protein binding; IPI.  
 DR InterPro: IPR006208; Cys\_knot.  
 DR InterPro: IPR006207; Cys\_knot\_C.  
 DR InterPro: IPR009041; PME\_inhibitor.  
 DR InterPro: IPR002919; TIL\_Cysrich.  
 DR InterPro: IPR006552; VC\_out.  
 DR InterPro: IPR001007; WVF\_C.  
 DR InterPro: IPR001846; WVF\_D.  
 DR Pfam: PF00007; Cys\_knot; 1.  
 DR Pfam: PF01826; TIL; 1.  
 DR Pfam: PF00093; wvc; 1.  
 DR Pfam: PF00094; wvd; 4.  
 DR SMART: SM00214; WVC; 6.  
 DR SMART: SM00215; WVC\_out; 4.

DR SMART: SM00216; WVD; 4.  
 DR PROSITE: PS01185; CTCK\_1; 1.  
 DR PROSITE: PS01225; CTCK\_2; 1.  
 DR PROSITE: PS01208; WFC\_1; 2.  
 DR PROSITE: PS0184; WFC\_2; 2.  
 KW Glycoprotein; Repeat; Signal; Polymorphism.  
 FT SIGNAL 1 25  
 FT CHAIN 26 5703 MUCIN 5B.  
 FT DOMAIN 77 225 WVD 1.  
 FT DOMAIN 329 386 TIL.  
 FT DOMAIN 426 580 WVD 2.  
 FT DOMAIN 858 918 WVD 1.  
 FT DOMAIN 896 1044 WVD 3.  
 FT DOMAIN 1457 1603 THR-RICH.  
 FT DOMAIN 1609 4873 THR-RICH.  
 FT DOMAIN 5005 5178 WVD 4.  
 FT DOMAIN 5353 5425 WFC 2.  
 FT DOMAIN 5462 5528 WFC 3.  
 FT DOMAIN 5594 5683 CTCK.  
 FT DISULFID 5594 5646 BY SIMILARITY.  
 FT DISULFID 5622 5660 BY SIMILARITY.  
 FT DISULFID 5626 5676 BY SIMILARITY.  
 FT DISULFID 5643 5678 BY SIMILARITY.  
 FT DISULFID 5645 5682 BY SIMILARITY.  
 FT CARBOHYD 145 145 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 201 201 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 254 254 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 402 402 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 516 516 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 806 806 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 930 930 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1277 1277 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1293 1293 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1557 1557 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1775 1775 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2192 2192 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2721 2721 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 3419 3419 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 3948 3948 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 4745 4745 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 4901 4901 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 4958 4958 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 4965 4965 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 4987 4987 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 5037 5037 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 5052 5052 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 5156 5156 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 5427 5427 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 5467 5467 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 5506 5506 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 5507 5507 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 5543 5543 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 5553 5553 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 5604 5604 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 5618 5618 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 5662 5662 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT VARIANT 5137 T -> S (in dbSNP:2672788).  
 FT CONFLICT 34 G -> E (IN REF. 2).  
 FT CONFLICT 95 FPGCN -> LPCLK (IN REF. 2).  
 FT CONFLICT 104 S -> C (IN REF. 2).  
 FT CONFLICT 142 E -> K (IN REF. 1).  
 FT CONFLICT 225 R -> S (IN REF. 2).  
 FT CONFLICT 330 PL -> T (IN REF. 2).  
 FT CONFLICT 337 E -> N (IN REF. 2).  
 FT CONFLICT 356 E -> K (IN REF. 2).  
 FT CONFLICT 362 G -> R (IN REF. 2).  
 FT CONFLICT 369 MISSING (IN REF. 2 AND 3).  
 FT CONFLICT 374 D -> N (IN REF. 2).  
 FT CONFLICT 393 RT -> TR (IN REF. 2).  
 FT CONFLICT 468 RK -> GR (IN REF. 2).  
 FT CONFLICT 512 L -> P (IN REF. 2).  
 FT CONFLICT 585 GAA -> AH (IN REF. 3).

```
FT CONFLICT 601 601 A -> S (IN REF. 3).
FT CONFLICT 629 629 DP -> RS (IN REF. 2).
FT CONFLICT 633 633 F -> L (IN REF. 2).
FT CONFLICT 676 676 A -> P (IN REF. 3).

Query Match 3.8%; Score 154.5; DB 1; Length 5703;
Best Local Similarity 20.2%; Pred. No. 1.2;
Matches 130; Conservative 58; Mismatches 281; Indels 173; Gaps 23;

Qy 102 SSDQATWQI-----TLPFKLGGMGGRGLAVDPNNDNLYFCAP 146
Dy |||||
Db |||||
Qy 147 SGKGLMRST-----DSGATQMTNPPDVGTVIANPTDTTGYQSDIQGVVWVAFD 196
Dy |||||
Db |||||
Qy 3008 SSKATSSSPRTATTLPLVLTATKSTATSTPTIPSSTLGTTGTSQNRPPHPMATMSTIH 3067
Dy |||||
Db |||||
Qy 197 KSSSLGQAKTIFV-----GVADNNPVFWRDGGATW-----QAVGAPGTGFIPIHK 245
Dy |||||
Db |||||
Qy 3068 PSSTPETHTSTVLTKATTTTRATSSMSTP---SSTPGTWTILTELTAATTAALPH-- 3122
Dy |||||
Db |||||
Qy 246 VFDPNVHLYIATNCGPYDSSGDYKFSVTSGTWTIRISPVSTDTANDYFGYGLTI 305
Dy |||||
Db |||||
Qy 3123 -----GTPSPGTWTILTEPSTTATVVTGSTATSS----- 3156
Dy |||||
Db |||||
Qy 306 DRQHPNTIMVATQISWPPDTIIFRSTDGGATWTRIDWTSYPNRSRLRYVLDISAEPMLTF 365
Dy |||||
Db |||||
Qy 3157 TRATAGTLKVLSTATPTPTVLISSRATSSSPGTA---TALP--ALRSTATTPTATSVT- 3209
Dy |||||
Db |||||
Qy 366 GVQNPVPVPGKLG--WDEMAIADFPNSDRMLYGTGATLYATNDLTKWDSGGQIHAPM 423
Dy |||||
Db |||||
Qy 3210 -----AIPSSLGATWRLSQTTP-----TATMSTATSPSTP---ETVHTSV 3250
Dy |||||
Db |||||
Qy 424 VKGLEETAVNDLISPPSGAPLISALDGLGTHADVTAPVS-----TFTSPVFTTGTSVD 479
Dy |||||
Db |||||
Qy 3251 LTTATTATRTGVSATPSTPGTATTKVPTTTTGTATPSSPGTALTTPVWISTTTTP 3310
Dy |||||
Db |||||
Qy 480 YAEINPSIIVRAGSFDPSQNDRHVA-----FSTDGKNWFGSGEPGGVT 525
Dy |||||
Db |||||
Qy 3311 -----TRGSTVTSSIPGTHTATVLTITTTTATGSMATPSSSTQTSQTPPSLT 3362
Dy |||||
Db |||||
Qy 526 TGGTVAASDGSREVFAPG-DPGQVYVAVGFGNSWAASQGVPAANAQIRSDRVNPKTFYA 584
Dy |||||
Db |||||
Qy 3363 TATTITATGTTNPSSTPGTTPPIPPVLTATTATPAATSSVTTPSSAL----- 3409
Dy |||||
Db |||||
Qy 585 LSGTFRSTDDGGVTFQVAPAGLPSSGAVGVFWFAVPGKEDGLWLAASGLYHSTNGSS 644
Dy |||||
Db |||||
Qy 3410 ---GTTHTPPVPNTTATHGRSLPSSP-----HTVP---TAWTSATSGILGTH--- 3453
Dy |||||
Db |||||
Qy 645 WSAITGVSSAVNVGFGKSAPSSYPVAVVGTIGGVTGAYBS 686
Dy |||||
Db |||||
Qy 3454 ---ITEPST-----GTSHTPAATTGTPFSPALSS 3481
Dy |||||
Db |||||

RESULT 6
APMU_PIG STANDARD; PRT; 1150 AA.
AC P12021;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Apomucin (Mucin core protein) (Fragment).
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Submaxillary gland;
RX MEDLINE=91236743; PubMed=2033060;
RA Eckhardt A.E., Timpte C.S., Abernethy J.L., Zhao Y., Hill R.L.;
RT "Porcine submaxillary mucin contains a cysteine-rich,
carboxyl-terminal domain in addition to a highly repetitive,
```

```
RT glycosylated domain.";
RL J. Biol. Chem. 266:9678-9686(1991).
RN [2]
RP SEQUENCE OF 1-503 FROM N.A.
RC TISSUE=Submaxillary gland;
RX MEDLINE=88087170; PubMed=2826455;
RA Timpte C.S., Eckhardt A.E., Abernethy J.L., Hill R.L.;
RT "Porcine submaxillary gland apomucin contains tandemly repeated,
identical sequences of 81 residues.";
RL J. Biol. Chem. 263:1081-1088(1988).
RN [3]
RP SEQUENCE OF 45-80.
RC TISSUE=Submaxillary gland;
RX MEDLINE=87280230; PubMed=3611111;
RA Eckhardt A.E., Timpte C.S., Abernethy J.L., Toumadje A.,
Johnson W.C. Jr., Hill R.L.;
RT "Structural properties of porcine submaxillary gland apomucin.";
RL J. Biol. Chem. 262:11339-11344(1987).
RN [4]
RP CARBOHYDRATE-LINKAGE SITES, AND SEQUENCE OF 45-125.
RC TISSUE=Submaxillary gland;
RX MEDLINE=97248516; PubMed=9092502;
RA Gerken T.A., Owens C.L., Pasumamthy M.;
RT "Determination of the site-specific O-glycosylation pattern of the
porcine submaxillary mucin tandem repeat glycopeptide. Model proposed
for the polypeptide:galnac transferase peptide binding site.";
RL J. Biol. Chem. 272:9709-9719(1997).
CC -|- FUNCTION: APOMUCIN IS PART OF MUCIN, THE MAJOR GLYCOPROTEIN
SYNTHESIZED AND SECRETED BY MUCOUS CELLS OF THE SUBMAXILLARY
GLAND. ITS HIGHLY VISCOUS AQUEOUS SOLUTIONS SERVE TO LUBRICATE
THE ORAL CAVITY AND TO PROTECT IT FROM THE EXTERNAL
ENVIRONMENT.
CC -|- SUBUNIT: INTERMOLECULAR DISULFIDE BONDS COULD HELP MAINTAIN A
MULTIMERIC MUCIN STRUCTURE.
CC -|- SUBCELLULAR LOCATION: Secreted.
CC -|- TISSUE SPECIFICITY: SUBMAXILLARY MUCOSAE.
CC -|- DOMAIN: CONTAINS TANDEMLY REPEATED, IDENTICAL SEQUENCES OF 81
RESIDUES.
CC -|- PTM: EXTENSIVELY O-LINKED GLYCOSYLATED ON SER AND THR RESIDUES OF
THE REPEAT UNITS. HIGHEST GLYCOSYLATION APPEARS TO OCCUR ON SER
RESIDUES WHICH HAVE GLY AT POSITIONS +2 OR -2 FROM THE
GLYCOSYLATION SITE OR, WHERE GLY IS THE PENULTIMATE RESIDUE. THE
PRESENCE OF PROLINE (USUALLY AT POSITION +3 OR -3) APPEARS TO ALSO
ENHANCE GLYCOSYLATION.
CC -|- SIMILARITY: Contains 1 WFVFC domain.
CC -|- SIMILARITY: Contains 1 C-terminal cysteine knot-like (CTCK) domain.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
DR EMBL; M61883; AAA30398.1; --
DR EMBL; M21174; AAA30990.1; --
DR InterPro; IPR006208; Cys knot.
DR InterPro; IPR006207; Cys knot_C.
DR InterPro; IPR001007; WFVFC.
DR Pfam; PF00007; Cys knot; 1.
DR SMART; SM00214; WVC; 1.
DR PROSITE; PS01185; CTCK_1; 1.
DR PROSITE; PS01225; CTCK_2; 1.
DR PROSITE; PS01208; WFVFC_1; 1.
DR PROSITE; PS01184; WFVFC_2; 1.
KW Glycoprotein; Repeat.
FT NON_TER 1
FT DOMAIN <1 368 81 AA TANDEM REPEATS.
FT REPEAT <1 44 1.
FT REPEAT 45 125 2.
FT REPEAT 126 206 3.
FT REPEAT 207 287 4.
```



CC	PRODUCE MULTIFORM BETA-AMYLASES AND A 48 kDa ALPHA-AMYLASE AFTER SECRETION.	
CC	-!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-alpha-glucosidic linkages in polysaccharides so as to remove successive maltose units from the non-reducing ends of the chains.	
CC	-!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic linkages in oligosaccharides and polysaccharides.	
CC	-!- SUBCELLULAR LOCATION: Secreted.	
CC	-!- SIMILARITY: In the N-terminal section; belongs to family 14 of glycosyl hydrolases.	
CC	-!- SIMILARITY: In the C-terminal section; belongs to family 13 of glycosyl hydrolases.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	
CC	-----	
CC	EMBL: M15817; AA085446.1; -.	
CC	EMBL: Y00150; CAA68344.1; -.	
CC	PIR: A29130; A29130.	
CC	HSSP: P36924; 1B9Z.	
CC	InterPro: IPR006589; Alp_amy1_cat sub.	
CC	InterPro: IPR006048; Alpha_amy1_C.	
CC	InterPro: IPR006047; Alpha_amy1_cat.	
CC	InterPro: IPR005085; CBM_25.	
CC	InterPro: IPR006046; Glyco_hydro_13.	
CC	InterPro: IPR001554; Glyco_hydro_14.	
CC	Pfam: PF00128; alpha-amy1ase; 1.	
CC	Pfam: PF02806; alpha-amy1ase_C; 1.	
CC	Pfam: PF03423; CBM_25; 2.	
CC	Pfam: PF01373; Glyco_hydro_14; 1.	
CC	PRINTS: PR00110; ALPHAAMYLASE.	
CC	PRINTS: PR00750; BETAAMYLASE.	
CC	SMART: SM00642; Aamy; 1.	
CC	SMART: SM00632; Aamy; C; 1.	
CC	PROSITE: PS00506; BETA_AMYLASE_1; 1.	
CC	PROSITE: PS00679; BETA_AMYLASE_2; 1.	
KW	Multifunctional enzyme; Hydrolase; Glycosidase; Signal;	
KW	Polysaccharide degradation; Repeat.	
FT	SIGNAL 1 35	
FT	CHAIN 36 1196	BETA/ALPHA-AMYLASE.
FT	DOMAIN 36 454	BETA-AMYLASE.
FT	REPEAT 455 558	
FT	REPEAT 565 668	
FT	DOMAIN 669 1196	ALPHA-AMYLASE.
FT	DISULFID 118 126	
FT	ACT_SITE 198 198	BY SIMILARITY.
FT	ACT_SITE 394 394	BY SIMILARITY.
FT	MUTAGEN 118 118	C->S: 5-FOLD DECREASE IN ACTIVITY.
FT	MUTAGEN 126 126	C->V: 20-FOLD DECREASE IN ACTIVITY.
FT	MUTAGEN 358 358	C->S: 60-FOLD DECREASE IN ACTIVITY.
FT	CONFLICT 1 1	M -> MIGL (IN REF. 3).
FT	CONFLICT 67 67	N -> S (IN REF. 3).
FT	CONFLICT 100 100	N -> D (IN REF. 3).
FT	CONFLICT 154 154	S -> N (IN REF. 3).
FT	CONFLICT 177 177	E -> Q (IN REF. 3).
FT	CONFLICT 227 228	NA -> KS (IN REF. 3).
FT	CONFLICT 330 330	G -> S (IN REF. 3).
FT	CONFLICT 425 425	N -> S (IN REF. 3).
FT	CONFLICT 493 493	D -> A (IN REF. 3).
FT	CONFLICT 532 532	S -> L (IN REF. 3).
FT	CONFLICT 559 559	A -> T (IN REF. 3).
FT	CONFLICT 665 665	A -> T (IN REF. 3).
FT	CONFLICT 681 681	D -> N (IN REF. 3).
FT	CONFLICT 686 686	T -> A (IN REF. 3).
FT	CONFLICT 725 728	AFTS -> VFSP (IN REF. 3).
FT	CONFLICT 736 736	N -> K (IN REF. 3).
FT	CONFLICT 741 741	N -> S (IN REF. 3).
FT	CONFLICT 758 758	S -> N (IN REF. 3).

OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 ON NCBI\_TaxID=10090;  
 RX MEDLINE=95104847; PubMed=7806222;  
 RA Walcz E., Deak F., Erhardt P., Coulter S.N., Fueloep C., Horvath P.,  
 RA Dooge K.J., Glant T.T.;  
 RT "Complete coding sequence, deduced primary structure, chromosomal  
 RT localization, and structural analysis of murine aggrecan.";  
 RL Genomics 22:364-371(1994).  
 RN [2]  
 RP SEQUENCE OF 211-326 FROM N.A.  
 RC STRAIN=129/SV;  
 RX MEDLINE=95004579; PubMed=7920633;  
 RA Watanabe H., Kimata K., Line S., Strong D., Gao L.-Y., Kozak C.A.,  
 RA Yamada Y.;  
 RT "Mouse cartilage matrix deficiency (cmd) caused by a 7 bp deletion in  
 RT the aggrecan gene.";  
 RL Nat. Genet. 7:154-157(1994).  
 RN [3]  
 RP INTERACTION WITH FBLN1  
 RX MEDLINE=99329059; PubMed=10400671;  
 RA Asberg A., Adam S., Kostka G., Timpl R., Heinigaard D.;  
 RA "Fibulin-1 is a ligand for the C-type lectin domains of aggrecan and  
 RT versican.";  
 RL J. Biol. Chem. 274:20444-20449(1999).  
 CC -!- FUNCTION: This proteoglycan is a major component of extracellular  
 CC matrix of cartilaginous tissues. A major function of this protein  
 CC is to resist compression in cartilage. It binds avidly to  
 CC hyaluronic acid via an amino-terminal globular region. May play a  
 CC regulatory role in the matrix assembly of the cartilage.  
 CC -!- SUBUNIT: Interacts with FBLN1.  
 CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix (By  
 CC similarity).  
 CC -!- DOMAIN: Two globular domains, G1 and G2, comprise the amino  
 CC terminus of the proteoglycan, while another globular region, G3,  
 CC makes up the COOH terminus. G1 contains link domains and thus  
 CC consists of three disulfide-bonded loop structures designated as  
 CC the A, B, B' motifs. G2 is similar to G1. The keratan sulfate (KS)  
 CC and the chondroitin sulfate (CS) attachment domains lie between G2  
 CC and G3.  
 CC -!- PTM: Contains mostly chondroitin sulfate, but also keratan sulfate  
 CC chains, N-linked and O-linked oligosaccharides.  
 CC -!- DISEASE: Defects in AGC1 are the cause of cartilage matrix  
 CC deficiency (CMD). CMD is an autosomal recessive syndrome  
 CC characterized by cleft palate, short limbs, tail and snout.  
 CC Mutation in strain CMD causes absence of aggrecan by truncation of  
 CC the protein (mutation in the G1 domain).  
 CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.  
 CC -!- SIMILARITY: Contains 4 link domains.  
 CC -!- SIMILARITY: Contains 1 EGF-like domain.  
 CC -!- SIMILARITY: Contains 1 C-type lectin family domain.  
 CC -!- SIMILARITY: Contains 1 Sushi (SCR) domain.  
 CC -!- SIMILARITY: Belongs to the aggrecan/versican proteoglycan family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L07049; AAC37670.1; -;  
 DR EMBL; S73722; AAB32160.1; -;  
 DR EMBL; S73721; AAB32160.1; JOINED.  
 DR PIR; A55182; A55182.  
 DR HSPF; P98066; ITSG.  
 DR MGD; MGI:99602; Agc1.  
 DR InterPro; IPR002353; AntifreezeII.

DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003006; Ig\_MHC.  
 DR InterPro; IPR003596; Ig\_V.  
 DR InterPro; IPR001304; Lectin\_C.  
 DR InterPro; IPR000538; Link\_  
 DR InterPro; IPR003324; SGXSG  
 DR InterPro; IPR000436; Sushi\_SCR\_CCP.  
 DR Pfam; PF00047; Ig; 1.  
 DR Pfam; PF00059; lectin\_c; 1.  
 DR Pfam; PF02339; SGXSG; 60.  
 DR Pfam; PF00084; sushi; 1.  
 DR Pfam; PF00193; Xlink; 4.  
 DR PRINTS; PR00356; ANTIFREEZEII.  
 DR PRINTS; PR01265; LINKMODULE.  
 DR ProDom; PD000918; Link; 4.  
 DR SMART; SM00032; CCP; 1.  
 DR SMART; SM00034; CLECT; 1.  
 DR SMART; SM00406; IGV; 1.  
 DR SMART; SM00445; LINK; 4.  
 DR PROSITE; PS00615; C-TYPE LECTIN 1; 1.  
 DR PROSITE; PS00041; C-TYPE LECTIN 2; 1.  
 DR PROSITE; PS00835; IG-LIKE; 1.  
 DR PROSITE; PS00290; IG-MHC; 1.  
 DR PROSITE; PS01241; LINK; 4.  
 KW Glycoprotein; Proteoglycan; Lectin; Signal; Sushi; Repeat;  
 KW Immunoglobulin domain.  
 FT SIGNAL 1 19 POTENTIAL.  
 FT CHAIN 20 2132 AGGECAN CORE PROTEIN.  
 FT DOMAIN 34 147 IG-LIKE V-TYPE.  
 FT DOMAIN 170 247 LINK 1.  
 FT DOMAIN 268 349 LINK 2.  
 FT DOMAIN 504 581 LINK 3.  
 FT DOMAIN 602 683 LINK 4.  
 FT DOMAIN 1918 2044 C-TYPE LECTIN.  
 FT DOMAIN 2048 2106 SUSHI.  
 FT DOMAIN 48 140 GI-A.  
 FT DOMAIN 152 247 GI-B.  
 FT DOMAIN 253 349 GI-B'.  
 FT DOMAIN 486 580 G2-B.  
 FT DOMAIN 587 682 G2-B'.  
 FT DOMAIN 685 803 KS.  
 FT DOMAIN 805 1231 CS-1.  
 FT DOMAIN 1232 1917 CS-2.  
 FT DOMAIN 1917 2132 G3.  
 FT DISULFID 51 133 BY SIMILARITY.  
 FT DISULFID 175 246 BY SIMILARITY.  
 FT DISULFID 199 220 BY SIMILARITY.  
 FT DISULFID 273 348 BY SIMILARITY.  
 FT DISULFID 297 318 BY SIMILARITY.  
 FT DISULFID 509 580 BY SIMILARITY.  
 FT DISULFID 533 554 BY SIMILARITY.  
 FT DISULFID 607 682 BY SIMILARITY.  
 FT DISULFID 631 652 BY SIMILARITY.  
 FT DISULFID 1922 1933 BY SIMILARITY.  
 FT DISULFID 1950 2042 BY SIMILARITY.  
 FT DISULFID 2018 2034 BY SIMILARITY.  
 FT DISULFID 2049 2092 BY SIMILARITY.  
 FT DISULFID 2078 2105 BY SIMILARITY.  
 FT CARBOHYD 126 126 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 239 239 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 333 333 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 387 387 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 611 611 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 667 667 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1675 1675 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT SITE 1171 1173 CELL ATTACHMENT SITE (POTENTIAL).  
 SQ SEQUENCE 2132 AA; 222008 MW; 0B2BCDF6C8DA163 CRC64;

Query Match 3.7%; Score 149; DB 1; Length 2132;  
 Best Local Similarity 19.5%; Pred. No. 0.74;  
 Matches 163; Conservative 96; Mismatches 286; Indels 292; Gaps 39;

QY 64 GYGVVSTAADPINTNKVAAVGMVTSWDPNDGAILRSSDQGA-----TWQITPL 114

Db 1105 GYVSGIPSGDGTET-----SASGVEDVSGLPSCGEGLETSASGVEDVGPSTROSLTSAS 1160  
Qy 115 PFKLGNNPGRGMRGLAVDPNDNILYFGAPSKGLWRSDTSQADWQSTQNFDP----- 169  
Db 1161 GVDVTGPPSGRGPETSVSGVGD-----FSGLPSCGEGLETSASGA--EDLSGLPSGKEDL 1215  
Qy 170 -----VGYIYANPT-DTTGYQSDIGQVVMVAFDKSSSLGQASKTIFVGVGA 214  
Db 1216 VGSASGALDFGKLPPLGSGQTPVNGPSPGSG-----EYSGADIGSPSS--GLP 1266  
Qy 215 DPNNPVWSDGGATWQAVGATGPTPHKGVFPVNVHLYIATNTGGPYDG----- 267  
Db 1267 D-----FSGLPSPGF-PTVSLVD--STLVEVITATTSSELEGRGTIGIS 1306  
Qy 268 SSGDVWKFVTSQWTRISVPSTDTANDYFGYSLGTLIDRQHPNTIMVATQISMPWTII 327  
Db 1307 GSGEV-----SGL-----PLGELDSD--ISGL-----PSGTSLGQASGSPDS-- 1344  
Qy 328 FRSTDGGATWRIWDMTYSYENRSLRYVLDISAEPWLTFGVQPNPFPSPKLGMMDEAMAI 387  
Db 1345 SGETSG-----PFVDSGQFPSSGVS-----EETSGI 1371  
Qy 388 DPFNSDRMLYGTATLYATNDLTKWD--SGQIHIAPMVKGL-----EET 430  
Db 1372 PEISGQPS--GTPDPTT-ATSCVTELNELSSQPDVSGDGSGLFGSGQSGITSVSGETS 1428  
Qy 431 AVNDLSPSPGAPL-----ISALDGLGFTHADYVAPSTIFT----- 468  
Db 1429 GISDLGQSPGPPVFGSGTATRTPLASGTISGSGESSGITFVDSFVEVPTTFRBEGL 1488  
Qy 469 -----SPVFTGTSDVYAEINLPSIIVRAGSFPDSSQPNDRHVAFSTDGKNWF 516  
Db 1489 GSVELSGFPSEGTSLGTSQTVDSQSSGALDSGLTSPTPE-----F 1532  
Qy 517 QGSEPG-----GVTGCTGAASA-DGSRFWAPGDP-----GQPVVY 552  
Db 1533 SGLPSGVAEVSGERGSGVETGSSLPAGFDSGLV--SGFPTVSLVDRTLVESITQAPTAQ 1590  
Qy 553 AVGCGNSW-----AASQGV-----ANQIRSDRVNPKTFVALSNGTFYR 592  
Db 1591 EAGEGSGILEFSGAHSGTPTDISGELSGSLDLSTLQSGQMETSTETSPSPY--FSGDPSS 1648  
Qy 593 STD--GGVTFQVAGLPSGAGVGMFHAVPCGKEDLMAASSGLVH----- 637  
Db 1649 TTDVSGESIAATGSGESSGLPEVTLNISELVEGTEPTVSOELCHGSPMTYISRLSEAS 1708  
Qy 638 --STNGSSWSAIT-----GVSSAV-----NVGFGKSAFGSPSYAVFVVGITGG 679  
Db 1709 GDASAGDLGAVTNPPGSGVEASVPEASSDLSAYPEAGVGVSAPEASSKLSFPDLHG 1768  
Qy 680 VTGAYRSDCGTT-----WVLINDQHQVGNWQQAITGDHANLRRVYIGTNG 726  
Db 1769 ITSAFHETDLEMTTPTSTEVNSNPWF--QEGTREGSAPEVSGESSTTSDIDTGTSG 1823

## RESULT 9

ISOA\_PSEAY STANDARD; PRT; 776 AA.  
AC P10342;  
DT 01-MAR-1989 (Rel. 10, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Isoamylase precursor (EC 3.2.1.68).  
GN IAM.  
OS Pseudomonas amyloclavosa.  
OC Bacteria; Proteobacteria.  
OX NCBI\_TaxID=32043;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SB-15;  
RX MEDLINE=88243808; PubMed=3379068;  
RA Amemura A., Chakraborty R., Fujita M., Nouni T., Futai M.;

"Cloning and nucleotide sequence of the isoamylase gene from Pseudomonas amyloclavosa SB-15.";  
J. Biol. Chem. 263:9271-9275(1988).  
[2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JD210;  
RX MEDLINE=91064385; PubMed=2248978;  
RA Chen J.H., Chen Z.Y., Chow T.Y., Chen J.C., Tan S.T., Hsu W.H.;  
"Nucleotide sequence and expression of the isoamylase gene from an isoamylase-hyperproducing mutant, Pseudomonas amyloclavosa JD210.";  
Biochim. Biophys. Acta 1087:309-315(1990).  
[3]  
RP SEQUENCE OF 744-776 FROM N.A.  
RC STRAIN=SB-15;  
RX MEDLINE=89327147; PubMed=2753857;  
RA Amemura A., Fujita M., Futai M.;  
"Transcription of the isoamylase gene (iam) in Pseudomonas amyloclavosa SB-15.";  
J. Bacteriol. 171:4320-4325(1989).  
[4]  
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).  
RX MEDLINE=98387895; PubMed=9719642;  
RA Katsuya Y., Mezaki Y., Kubota M., Matsuura Y.;  
"Three-dimensional structure of Pseudomonas isoamylase at 2.2-A resolution.";  
J. Mol. Biol. 281:885-897(1998).  
CC -I- CATALYTIC ACTIVITY: Hydrolysis of alpha-(1,6)-D-glucosidic branch linkages in glycogen, amylopectin and their beta-limits dextrans.  
CC -I- COPACITOR: Binds 1 calcium ion per subunit.  
CC -I- SUBUNIT: Monomer.  
CC -I- INDUCTION: By maltose.  
CC -I- SIMILARITY: Belongs to family 13 of glycosyl hydrolases.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
EMBL: J03871; AAA25854.1; -;  
EMBL: X13378; CAA31754.1; -;  
PDB: 1BF2; 12-AUG-98.  
InterPro: IPR006047; Alpha amyl\_cat.  
InterPro: IPR004193; Glyco\_hydro\_13N.  
InterPro: IPR007110; Ig-like.  
Pfam: PF00128; alpha-amylase; 1.  
Pfam: PF02922; isoamylase N; 1.  
Hydrolase; Glycosidase; Calcium-binding; Signal; 3D-structure.  
SIGNAL 1 26  
CHAIN 27 776 ISOAMYLASE.  
FT ACT SITE 401 401  
FT ACT SITE 461 461  
FT ACT SITE 536 536  
FT METAL 154 154  
FT METAL 255 255  
FT METAL 256 256  
FT METAL 258 258  
FT METAL 285 285  
FT DISULFID 410 422  
FT DISULFID 546 616  
FT DISULFID 738 766  
FT CONFLICT 8 8  
FT CONFLICT 126 126  
FT CONFLICT 169 171  
FT CONFLICT 386 386  
FT CONFLICT 413 416  
FT CONFLICT 454 489  
A -> G (IN REF. 1).  
F -> C (IN REF. 1).  
GAS -> AH (IN REF. 1).  
L -> V (IN REF. 1).  
GAYT -> AVH (IN REF. 1).  
SGLDLFAEPWAGTSGYQLGGPQGWSEWNGLPDS ->  
TVWICLNLGPPSAATRTSWDSRRVVRVSVPRQ (IN  
REF. 1).  
WP -> S (IN REF. 1).  
APRKAHPA -> RSARHIP (IN REF. 1).  
CONFLICT 555 556  
CONFLICT 650 657

FT	TURN	30	31	FT	HELIX	437	444
FT	STRAND	34	37	FT	STRAND	447	447
FT	TURN	39	40	FT	TURN	450	451
FT	STRAND	43	49	FT	STRAND	456	460
FT	STRAND	55	61	FT	TURN	467	468
FT	STRAND	71	74	FT	TURN	472	473
FT	STRAND	76	77	FT	TURN	477	478
FT	TURN	80	81	FT	STRAND	480	482
FT	STRAND	82	88	FT	HELIX	484	495
FT	HELIX	89	94	FT	STRAND	497	497
FT	TURN	95	96	FT	TURN	498	499
FT	STRAND	102	109	FT	STRAND	500	500
FT	TURN	110	111	FT	HELIX	504	511
FT	STRAND	112	112	FT	TURN	512	513
FT	TURN	116	117	FT	HELIX	515	518
FT	TURN	120	121	FT	HELIX	519	521
FT	TURN	124	125	FT	TURN	522	522
FT	TURN	132	133	FT	HELIX	525	527
FT	TURN	139	140	FT	STRAND	528	530
FT	STRAND	143	143	FT	HELIX	540	543
FT	TURN	146	147	FT	TURN	544	545
FT	STRAND	151	151	FT	TURN	555	556
FT	HELIX	163	166	FT	TURN	572	577
FT	STRAND	167	167	FT	HELIX	581	597
FT	TURN	170	174	FT	STRAND	601	605
FT	TURN	178	180	FT	TURN	606	607
FT	STRAND	184	185	FT	HELIX	608	610
FT	HELIX	202	204	FT	TURN	611	611
FT	STRAND	207	210	FT	TURN	615	616
FT	HELIX	212	216	FT	TURN	621	622
FT	TURN	217	218	FT	TURN	626	628
FT	TURN	220	221	FT	HELIX	636	654
FT	HELIX	224	226	FT	HELIX	656	658
FT	TURN	227	228	FT	TURN	666	668
FT	HELIX	230	235	FT	STRAND	669	672
FT	TURN	236	236	FT	TURN	674	675
FT	HELIX	237	243	FT	STRAND	678	678
FT	TURN	244	244	FT	HELIX	681	684
FT	STRAND	247	250	FT	TURN	685	685
FT	STRAND	254	254				
FT	STRAND	256	256				
FT	TURN	259	262				
FT	TURN	265	266				
FT	TURN	269	270				
FT	STRAND	278	278				
FT	STRAND	284	284				
FT	HELIX	286	288				
FT	TURN	294	295				
FT	HELIX	296	310				
FT	TURN	311	312				
FT	STRAND	314	319				
FT	TURN	325	326				
FT	STRAND	329	329				
FT	STRAND	338	339				
FT	HELIX	343	351				
FT	STRAND	352	352				
FT	STRAND	354	354				
FT	TURN	356	357				
FT	STRAND	361	361				
FT	STRAND	371	371				
FT	TURN	373	374				
FT	HELIX	376	391				
FT	TURN	392	392				
FT	STRAND	397	400				
FT	TURN	401	402				
FT	HELIX	403	407				
FT	STRAND	408	408				
FT	TURN	417	418				
FT	TURN	420	421				
FT	TURN	423	424				
FT	STRAND	429	429				
FT	TURN	431	432				
FT	TURN	434	435				

Query Match 3.7%; Score 148.5; DB 1; Length 776;  
Best Local Similarity 20.7%; Pred. No. 0.24;  
Matches 171; Conservative 88; Mismatches 275; Indels 291; Gaps 47;

QY	27	GAPGILYVRTDIGMYRWDAANGRWIPLLDWGVNNWNGVVSIAADPINTKVAWAVG	86
Db	96	GITGAVY-----YGYRAWGNPNWPYASNWKGSGQAGFVSDVDANGDRFNPKLLDDPY	147
QY	87	MYTNSWDPNDGAILRSSDQCATWQITPLPFKLGNNPGRGMGERLAVDPNNNDILYFEGAP	146
Db	148	AEVSDQP-----LNPSNQ-----NGNVFASGASYR-----TTDSGIY--AP	182
QY	147	SKGLWRSTDSGATWSQMTNFPDV--GTYIANPTD--TTGYQSDIQGVVWVAFDKSS--SS	201
Db	183	KGVWLVPSQTGTGKPTRAQKDDVIYEVHVRGTEQDTSIPAQYRGTYTYGAGLKASYLAS	242
QY	202	LCQASKTIFVGAD-----PN---NPVEW-----SRDGGATWQAVCAPTG-F	240
Db	243	LG-VTAVEFLPVQETQNDANDVVPNSDANQNYGYMTENYFSPDRRYAYNKAAGGPTAEF	301
QY	241	IPHKGVDFPNHVLVIATSNCTGGPYDGSDDVWKFVSVTSGTWTRISFVPST-----	291
Db	302	QMVQAFHNAGIKVYMDV-----VYNHTAEGGTWTSSTPTTATISWRGLDN	348
QY	292	-----DTANDYF-GYSGLTIDRQHNTI---MVATOISWVPTIIFRSTGGATWTRIW	341
Db	349	ATYYELTSGNQYFYDNTGTIGANFNTYNTVAQNLIIVDSLAYWANTM---GVDGFR----	400
QY	342	DWTS-YPNRSLRYVLDISAEPWLTFCVQPNPPVPSKLGWMDAIDAID-----PNSDRML	396
Db	401	DLASVLGNSCLNGAYTASA-----PNCPCNGGYNFDAADSNVAINRLREFTVRPAA	451
QY	397	YGTGATLYATNDLTKWDSGGQIH-IAPMVKGLLEE-----TAVNDLISPPSGAPLI	445

```

Db 452 GSGGLDLFA---EPWAIGNSYQLGFGPQGWSEWNGLFRDLSLROAQNEL-----GSMTI 502
Qy 446 SALGDLGGTHADVAVPSTFTSPVFTTGTSDVYAEINPSIIIV-----AGS 493
Db 503 YVTQDANDFGS-----SNLFQSGSRPWNINFDVHGMKLVKOVYSCNGANNSQAWP 556
Qy 494 FDPSSQPNDRHVAFTDGGK-----NWPQSGPBGVTTGGTV---AASADGSRF-VWAPGD 545
Db 557 YGPS-----DGGTSTNYSWQDMSAG---TGAADVQDQRAARTGMAPEMLSAGT 601
Qy 546 P-----GQPVVYAVGFGN-----SWAASQGVPAQAQIRSDR-----V 577
Db 602 PLMQGDEYLRTLQCNNAYNLDSSANWLITYSWTTDQ---SNFYTFQAQLIAFRKAHPAL 658
Qy 578 NPKTFYALNSGTFYRSTDDGVTGPVAAGLPSSGAVG-----VNFHAPVPGKGD 626
Db 659 RPSSWYSGSLTWY-----QP-----SGAVADSNYNWNTSNYAIAVAINGPS-- 700
Qy 627 LWLAASSGLYHSTNG-----GSSWSAIT-----GVSSAVNVGFGKSAPG 665
Db 701 --LGDSNSIYVAYNGWSSSVFTLPAPPSGTQWYRVTTDCMWDGASTFV-----APG 751
Qy 666 SSYPVAVFVGTIGVGTGAYRSDDCGTTWVLINDDQHQYGNWQAI 710
Db 752 SE-----TLIGG-----AGTT-----YQCGQSL 770

RESULT 10
ISOA PSESP
ID ISOA PSESP STANDARD; PRT; 776 AA.
AC P26501;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Isoamylase precursor (EC 3.2.1.68).
GN IAM.
OS Pseudomonas sp. (strain SMP1).
OC Bacteria; Proteobacteria.
OK NCBI_TaxID=306;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 27-44.
RX MEDLINE=89381677; PubMed=278432;
RA Tognoni A., Carrera P., Galli G., Lucchese G., Camerini B.,
RA Grandi G.;
RT "Cloning and nucleotide sequence of the isoamylase gene from a strain
RT of Pseudomonas sp.;"
RL J. Gen. Microbiol. 135:37-45(1989).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of alpha-(1,6)-D-glucosidic branch
CC linkages in glycogen, amylopectin and their beta-limits dextrins.
CC -1- COPACTOR: Binds 1 calcium ion per subunit (By similarity).
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- INDUCTION: By maltose.
CC -1- SIMILARITY: Belongs to family 13 of glycosyl hydrolases.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; M25247; AAA25855.1; -.
CC EMBL; A10909; CAA00929.1; -.
CC HSPG; P10342; IBF2.
CC InterPro; IPR006047; Alpha amyl cat.
CC InterPro; IPR004193; Glyco_hydro_13N.
CC InterPro; IPR007110; IG-like.
CC Pfam; PF00128; alpha-amylase; 1.
CC Pfam; PF02922; isoamylase N; 1.
CC Hydrolase; Glycosidase; Calcium-binding; Signal.
CC SIGNAL
1 26

```

```

FT CHAIN 27 776 ISOAMYLASE.
FT ACT SITE 401 401 BY SIMILARITY.
FT ACT SITE 461 461 BY SIMILARITY.
FT ACT SITE 536 536 BY SIMILARITY.
FT METAL 154 154 CALCIUM (BY SIMILARITY).
FT METAL 255 255 CALCIUM (BY SIMILARITY).
FT METAL 256 256 CALCIUM (VIA CARBONYL OXYGEN) (BY
SIMILARITY).
FT METAL 258 258 CALCIUM (BY SIMILARITY).
FT METAL 285 285 CALCIUM (BY SIMILARITY).
FT DISULFID 410 422 BY SIMILARITY.
FT DISULFID 546 616 BY SIMILARITY.
FT DISULFID 738 766 BY SIMILARITY.
SQ SEQUENCE 776 AA; 83656 MW; A5E4C02EF026A3A4 CRC64;

Query Match 3.7%; Score 148.5; DB 1; Length 776;
Best Local Similarity 20.7%; Pred. No. 0.24; Indels 291; Gaps 47;
Matches 171; Conservative 88; Mismatches 275;

Qy 27 GAPGILYVRTDIGMYRMDAANGRIPLLDVGMNMGYNGVVSIAADPINTNKVMAAVG 86
Db 96 GITGAVY-----YGYRANGPNWPYASNMGKSGQAGFVSDVDANGDRFNENKLLDPY 147
Qy 87 MYTNSWDPNDGAILRSSDQAGATQTPLPFLKGMNPGRGGERLAVDPNNDNILYFGAP 146
Db 148 AQEVSQDP-----LNPSNQ-----NGNVFASGASYR-----TTDSGIY--AP 182
Qy 147 SGKGLWRSTDSGATWSQMTNFPDV---GTIANPTD-TTGYQSDIOGVVWVAFDKS--SS 201
Db 183 KGVVLVPSTQSTGTKPTRAQKDDVIYVHVHVGFTQEDTSIPAQYRGTYVYAGLKAASYLAS 242
Qy 202 LQOAKTITFVGAD-----PN---NPVFW-----SRDGGATQWAVGAPTG-F 240
Db 243 LG-VTAVEFLPVQETQNDANDVVPNSDANQNYGMNTENYFSPDRRYAYNKAGGTAEF 301
Qy 241 IPHKGVPDPVNVHLYIATSTGGPYDGGSDGVWKFSTVSTGTWTRISPVST----- 291
Db 302 QAMVQAAPHNAGIKVYMDV-----VYNHTAEGGTWTSDDPTTIYSWRGLDN 348
Qy 292 -----DTANDYF-GYSGLTITDRQHPNTI---MVATQISWPDPTIIFRSTDGATWTRIW 341
Db 349 TTYTYELTSGNQFYDNTGIGANFNFTNTVAQNLI VDSLAYWANTM---GVDGFR-----F 400
Qy 342 DWTS-YENRSLRYVLDISAEPLWTFGVQPNPVPSPKLGWMDMAID----PFNSDRML 396
Db 401 DLASVLGNSCLNGAYTASA-----PNCNGGYNFDDAADSNAVAINRILREFTVRPAA 451
Qy 397 YGTGATLYATNDLTWKDSCGQIH-IAPMVYKGLEE-----TAVNDLISPPSGAPLI 445
Db 452 GSGGLDLFA---EPWAIGNSYQLGFGPQGWSEWNGLFRDLSLROAQNEL-----GSMTI 502
Qy 446 SALGDLGGTHADVAVPSTFTSPVFTTGTSDVYAEINPSIIIV-----AGS 493
Db 503 YVTQDANDFGS-----SNLFQSGSRPWNINFDVHGMKLVKOVYSCNGANNSQAWP 556
Qy 494 FDPSSQPNDRHVAFTDGGK-----NWPQSGPBGVTTGGTV---AASADGSRF-VWAPGD 545
Db 557 YGPS-----DGGTSTNYSWQDMSAG---TGAADVQDQRAARTGMAPEMLSAGT 601
Qy 546 P-----GQPVVYAVGFGN-----SWAASQGVPAQAQIRSDR-----V 577
Db 602 PLMQGDEYLRTLQCNNAYNLDSSANWLITYSWTTDQ---SNFYTFQAQLIAFRKAHPAL 658
Qy 578 NPKTFYALNSGTFYRSTDDGVTGPVAAGLPSSGAVG-----VNFHAPVPGKGD 626
Db 659 RPSSWYSGSLTWY-----QP-----SGAVADSNYNWNTSNYAIAVAINGPS-- 700
Qy 627 LWLAASSGLYHSTNG-----GSSWSAIT-----GVSSAVNVGFGKSAPG 665
Db 701 --LGDSNSIYVAYNGWSSSVFTLPAPPSGTQWYRVTTDCMWDGASTFV-----APG 751
Qy 666 SSYPVAVFVGTIGVGTGAYRSDDCGTTWVLINDDQHQYGNWQAI 710
Db 752 SE-----TLIGG-----AGTT-----YQCGQSL 770

```



Db 752 SE-----TLIGG-----AGTT-----YQCGQSL 770

RESULT 11

BICA\_SALTY  
ID BICA\_SALTY STANDARD; PRT; 1953 AA.  
AC P25927; P25928; Q9XCQ3;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Putative surface-exposed virulence protein bigA precursor.  
GN BIGA OR STM3478.  
OS Salmonella typhimurium.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Salmonella.  
OX NCBI\_TaxID=602;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 14028;  
RA Stojiljkovic I., Valentine P., Heffron F.;  
RT "Salmonella typhimurium rhs homolog."  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;  
RX MEDLINE=21534948; PubMed=11677609;  
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
RA Courtney L., Porwollik S., Ali J., Dante M., Du J., Hou S., Layman D.,  
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
RA Waterston K., Wilson R.K.;  
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
LT2."  
RL Nature 413:852-856(2001).  
RN [3]  
RP SEQUENCE OF 1-765 FROM N.A.  
RC STRAIN=LT2;  
RX MEDLINE=91100301; PubMed=1987123;  
RA Wu J.Y., Siegel L.M., Kredich N.M.;  
RT "High-level expression of Escherichia coli NADPH-sulfite reductase:  
requirement for a cloned *cydG* plasmid to overcome limiting shiroheme  
cofactor."  
RT J. Bacteriol. 173:325-333(1991).  
RL J. Bacteriol. 173:325-333(1991).  
CC -! CAUTION: Ref.3 sequence differs from that shown due to frameshifts  
in positions 414 and 732.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; AF133696; AAD39458.1; -;  
DR EMBL; AE008859; AAL22340.1; -;  
DR EMBL; M64506; AAR27042.1; ALT\_FRAME.  
DR EMBL; M64506; AAR27043.1; ALT\_FRAME.  
DR PIR; C39200; C39200.  
DR PIR; D39200; D39200.  
DR StyGene; SG10437; bigA.  
KW Virulence; Repeat; Signal; Complete proteome.  
FT SIGNAL 1 27 POTENTIAL.  
FT CHAIN 28 1953 PUTATIVE SURFACE-EXPOSED VIRULENCE  
FT DOMAIN 101 252 PROTEIN BIGA.  
FT REPEAT 101 103 15 X 11 AA TANDEM REPEATS.  
FT REPEAT 104 113 1 (INCOMPLETE).  
FT REPEAT 114 122 2 (INCOMPLETE).  
FT REPEAT 123 133 3 (INCOMPLETE).  
FT REPEAT 124 144 4.  
FT REPEAT 134 144 5.  
FT REPEAT 145 155 6.  
FT REPEAT 156 166 7.

RESULT 12

FT	REPEAT	167	177	8.
FT	REPEAT	178	188	9.
FT	REPEAT	189	199	10.
FT	REPEAT	200	210	11.
FT	REPEAT	211	221	12.
FT	REPEAT	222	232	13.
FT	REPEAT	233	243	14.
FT	REPEAT	244	252	15 (INCOMPLETE).
FT	CONFLICT	207	207	D -> DRGDDVTPDD (IN REF. 1).
FT	CONFLICT	514	514	A -> R (IN REF. 3).
FT	CONFLICT	1698	1698	D -> N (IN REF. 1).
FT	CONFLICT	1795	1798	OYLE -> ITLO (IN REF. 1).
FT	CONFLICT	1836	1837	SA -> T (IN REF. 1).
SEQ	SEQUENCE	1953 AA;	200150 MW;	611B3F1C954D91AE CRC64;

Query Match 3.7%; Score 148.5; DB 1; Length 1953;  
Best Local Similarity 20.5%; Pred. No. 0.71;  
Matches 158; Conservative .83; Mismatches 274; Indels 257; Gaps 34;

QY	44	WDAANG-----RWIPLLDVGMNNGVYVSIADPINTNKVMAAVGMYTNMSD---P	94
Db	1344	WNTATGVINGINPDAPLINLGRYNGFNQGTINVOGD---NAV--AISGTTSSYVNLV	1397
QY	95	NDGAILRSSDQATWQITPLPKLGNMGP-----RGMGERLAVDPNNDNLIYFGAPSGK	150
Db	1398	NSGTTNVGTEQK-----EDGTNGTGLIGKGNATTINNTADGVINNVYADDSYA	1448
QY	151	LWRSDSGATWQMTNFPDVGTYIANPTDTTGQSDIQGVWVAFDKSSSLGQASKTIF	210
Db	1449	FGKTKAIINNGEINLLCDSCGDIYAP-GTTGTQNDHNGTADIVIDAITAPTEGS--IP	1505
QY	211	VGVADEPNFVWSRDGATWQAVPGAPTGFPHKGVDFPNVHLYIATNTGGPYDGSSG	270
Db	1506	TPPADENAPQQLSN-----DVTMTKNADVATDSVS-----YIVGTNA-----	1531
QY	271	DVWKFSV-----TSGTWTTRISVPSTDTTANDYFGYGLIDHROHPNTIMVATQ	318
Db	1532	TLKANLVIQDNVQVDTGTGTADTTVVVDNAFTGSNIQAGDNIT-----STS	1580
QY	319	ISWMPDITFRSTDGATWTRIDWTSYPNRSRLRYVLDISAEPLWTFGQPNPVPSPKL	378
Db	1581	VVM--NAQSQDAGNV-----DVTMTKNADVATDSVS-----	1614
QY	379	GMMDAMADIPNSDRMLY-----GTGATLYATNDLTWKDMSGQ-----	417
Db	1615	---DVAQALDAGYTNNELYTSLVGTAEI---NSALKQVSGAQATTVFPEARVLNRF	1668
QY	418	---IHIAPMVK-GLIETAVNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSPVFT	474
Db	1669	MLADAAPQIKDGLAFNVV-----AKGDP-----RAELGNDTQYDMLALRQTLDLTA--SQ	1716
QY	475	GTSVDVYAEINPISIIVRAGSFPDSSQPNDRHVAFTDGGKNWFGQSPGGVTTGGTVAASA	534
Db	1717	NUTLEY-----GIARLDG	1729
QY	535	DGSRFVWAPDPGQPVVYAVGFG---NSWAASQGVPAQAQIRSVNPKTFYALSNGTFYR	592
Db	1730	DGSK---TAGDNLTCGYSQFGLKXSMAPDEGLAWNLSRLYDVHNLDSRSRYAGDVNK	1786
QY	593	STD-----GGVTFPQVAAGLPSSGAVGYMF-HAVPG---KE---GDLWLAASS	633
Db	1787	IADSDMRQYLEFRSEGAKTFTWMDALKVTPYAGVKFRHTMEDGYKERSAGDFNLMSNS	1846
QY	634	-----GLYHSTNGSSWSAITGVSSAVNVGFGKS-----APGSSYPAPVVG	675
Db	1847	GNETAVDSTVGLKLDYAGKDGWSATATLEGGNLSYSKQRTASLQGAAGQSF-GVDDGQ	1905
QY	676	TIGGYTG-----AYRSDCGTTTWVLINDDQHQYGNWGOAITGDHANLRVY	721
Db	1906	KGGVNGLATIGVKYSSND-----TALHLDAQWKEDGISDKGFMUNVKTF	1952

PGCA\_CHICK  
 ID PGCA\_CHICK STANDARD; PRT; 2109 AA.  
 AC P07898; Q90810; Q90820; Q90991; Q91047;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Aggrecan core protein precursor (Cartilage-specific proteoglycan core  
 DE protein) (CSPCP).  
 GN AGC1.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 (1)  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=White leghorn; TISSUE=Embryo;  
 RX MEDLINE=94043149; PubMed=8226878;  
 RA Li H., Schwartz N.B., Vertel B.M.;  
 RT "cDNA cloning of chick cartilage chondroitin sulfate (aggrecan) core  
 RT protein and identification of a stop codon in the aggrecan gene  
 RT associated with the chondrodystrophy, nanomelia.";  
 RL J. Biol. Chem. 268:23504-23511(1993).  
 (2)  
 RN SEQUENCE OF 1042-1559 FROM N.A.  
 RP TISSUE=Embryo;  
 RX MEDLINE=90307744; PubMed=1694853;  
 RA Krueger R.C. Jr., Fields T.A., Mensch J.R. Jr., Schwartz N.B.;  
 RT "Chick cartilage chondroitin sulfate proteoglycan core protein. II.  
 RT Nucleotide sequence of cDNA clone and localization of the S103L  
 RT epitope.";  
 RL J. Biol. Chem. 265:12088-12097(1990).  
 (3)  
 RN SEQUENCE OF 1-1855 AND 1893-2109 FROM N.A.  
 RP TISSUE=Cartilage;  
 RX MEDLINE=93111968; PubMed=1339285;  
 RA Chandrasekaran L., Tanzer M.L.;  
 RT "Molecular cloning of chicken aggrecan. Structural analyses.";  
 RL Biochem. J. 288:903-910(1992).  
 (4)  
 RN ERRATUM.  
 RP MEDLINE=94107258; PubMed=8280087;  
 RX Chandrasekaran L., Tanzer M.L.;  
 RL Biochem. J. 296:885-887(1993).  
 (5)  
 RN SEQUENCE OF 1492-1610 FROM N.A.  
 RP STRAIN=White leghorn; TISSUE=Chondrocytes;  
 RX MEDLINE=95128519; PubMed=7827752;  
 RA Primorac D., Stover M.L., Clark S.H., Rowe D.W.;  
 RT "Molecular basis of nanomelia, a heritable chondrodystrophy of  
 RT chicken.";  
 RL Matrix Biol. 14:297-305(1994).  
 (6)  
 RN SEQUENCE OF 1894-2109 FROM N.A.  
 RP MEDLINE=8908500; PubMed=3170613;  
 RA Tanaka T., Har-El R., Tanzer M.L.;  
 RT "Partial structure of the gene for chicken cartilage proteoglycan  
 RT core protein.";  
 RL J. Biol. Chem. 263:15831-15835(1988).  
 (7)  
 RN SEQUENCE OF 1693-1855 AND 1893-2109 FROM N.A.  
 RP MEDLINE=86259736; PubMed=3460082;  
 RX Sai S., Tanaka T., Koshier R.A., Tanzer M.L.;  
 RT "Cloning and sequence analysis of a partial cDNA for chicken  
 RT cartilage proteoglycan core protein.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 83:5081-5085(1986).  
 -!- FUNCTION: This proteoglycan is a major component of extracellular  
 CC matrix of cartilaginous tissues. A major function of this protein  
 CC is to resist compression in cartilage. It binds avidly to  
 CC hyaluronic acid via an amino-terminal globular region. May play a  
 CC regulatory role in the matrix assembly of the cartilage.  
 CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix (By  
 CC similarity).

-!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=1;  
 CC IsoId=P07898-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=P07898-2; Sequence=VSP\_003073;  
 CC -!- DOMAIN: Two globular domains, G1 and G2, comprise the amino  
 CC terminus of the proteoglycan, while another globular region, G3,  
 CC makes up the COOH terminus. G1 contains link domains and thus  
 CC consists of three disulfide-bonded loop structures designated as  
 CC the A, B, B' motifs. G2 is similar to G1. The keratan sulfate (KS)  
 CC and the chondroitin sulfate (CS) attachment domains lie between G2  
 CC and G3.  
 CC -!- PTM: Contains mostly chondroitin sulfate, but also keratan sulfate  
 CC chains, N-linked and O-linked oligosaccharides.  
 CC -!- DISEASE: DEFECTS IN AGC1 ARE THE CAUSE OF NANOMELIA, A LETHAL  
 CC CONNECTIVE TISSUE DISORDER AFFECTING CARTILAGE DEVELOPMENT  
 CC (CHONDRODYSPLASIA) CHARACTERIZED BY SHORTENED AND MALFORMED LIMBS.  
 CC AGGREGAN IS TRUNCATED AT ITS C-TERMINAL IN THE CS-2 BINDING DOMAIN  
 CC AND IS NOT ANYMORE SECRETED FROM THE CHONDROCYTES.  
 CC -!- SIMILARITY: Belongs to the aggrecan/versican proteoglycan family.  
 CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.  
 CC -!- SIMILARITY: Contains 4 link domains.  
 CC -!- SIMILARITY: Contains 1 EGF-like domain.  
 CC -!- SIMILARITY: Contains 1 C-type lectin family domain.  
 CC -!- SIMILARITY: Contains 1 Sushi (SCR) domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; L21913; AAB19128.1; -  
 CC EMBL; M38187; AAA48731.1; -  
 CC EMBL; M88101; -; NOT ANNOTATED\_CDS.  
 CC EMBL; S74657; AAC60751.1; -  
 CC EMBL; S74656; AAC60751.1; JOINED.  
 CC EMBL; J04028; AAA48719.1; -  
 CC EMBL; M13993; AAA48720.1; -  
 CC PIR; I50421; I50421.  
 CC HSP; P08709; 1BF9.  
 CC InterPro; IPR002353; AntifreezeII.  
 CC InterPro; IPR000152; Asx\_hydroxyl\_S.  
 CC InterPro; IPR000742; EGF 2.  
 CC InterPro; IPR001881; EGF Ca.  
 CC InterPro; IPR006209; EGF-like.  
 CC InterPro; IPR007110; Ig-like.  
 CC InterPro; IPR003599; Ig.  
 CC InterPro; IPR001304; Lectin C.  
 CC InterPro; IPR000538; Link.  
 CC InterPro; IPR003324; SGXSG.  
 CC InterPro; IPR000436; Sushi\_SCR\_CCP.  
 CC Pfam; PF00008; EGF; 1.  
 CC Pfam; PF00047; ig; 1.  
 CC Pfam; PF00059; lectin c; 1.  
 CC Pfam; PF02339; SGXSG; 56.  
 CC Pfam; PF00084; sushi; 1.  
 CC Pfam; PF00193; Xlink; 4.  
 CC PRINTS; PR00356; ANTIFREEZEII.  
 CC PRINTS; PR01265; LINKMODULE.  
 CC ProDom; PD000918; Link; 4.  
 CC SMART; SM00032; CCP; 1.  
 CC SMART; SM00034; CLECT; 1.  
 CC SMART; SM00179; EGF CA; 1.  
 CC SMART; SM00409; IG; 1.  
 CC SMART; SM00445; LINK; 4.  
 CC PROSITE; PS00010; ASX\_HYDROXYL; 1.  
 CC PROSITE; PS00615; C-TYPE LECTIN 1; 1.  
 CC PROSITE; PS00041; C-TYPE LECTIN 2; 1.  
 CC PROSITE; PS00022; EGF\_1; 1.

```

DR PROSITE; PS00026; EGF 3; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS01835; IG_LIKE; 1.
DR PROSITE; PS01241; LINK; 4.
KW Glycoprotein; Proteoglycan; Lectin; Signal; Sushi; EGF-like domain;
KW Alternative splicing; Repeat; Immunoglobulin domain.
FT SIGNAL 1 16
FT CHAIN 17 2109
FT DOMAIN 34 143
FT DOMAIN 166 243
FT DOMAIN 264 346
FT DOMAIN 537 614
FT DOMAIN 635 716
FT DOMAIN 1363 1742
FT DOMAIN 1855 1892
FT DOMAIN 1901 2019
FT DOMAIN 2023 2081
FT DOMAIN 48 137
FT DOMAIN 148 243
FT DOMAIN 249 346
FT DOMAIN 519 613
FT DOMAIN 620 715
FT DOMAIN 718 803
FT DOMAIN 805 1264
FT DOMAIN 1265 1742
FT DOMAIN 1893 2109
FT DISULFID 51 129
FT DISULFID 171 242
FT DISULFID 195 216
FT DISULFID 269 345
FT DISULFID 293 314
FT DISULFID 542 613
FT DISULFID 566 587
FT DISULFID 640 715
FT DISULFID 664 685
FT DISULFID 1859 1870
FT DISULFID 1864 1879
FT DISULFID 1881 1890
FT DISULFID 1897 1908
FT DISULFID 1925 2017
FT DISULFID 1933 2009
FT DISULFID 2024 2067
FT DISULFID 2053 2080
FT CARBOHYD 76 76
FT CARBOHYD 122 122
FT CARBOHYD 330 330
FT CARBOHYD 388 388
FT CARBOHYD 439 439
FT CARBOHYD 644 644
FT CARBOHYD 700 700
FT CARBOHYD 765 765
FT CARBOHYD 801 801
FT VARSPIC 1856 1892
FT CONFLICT 362 362

Query Match 3.6%; Score 146; DB 1; Length 2109;
Best Local Similarity 21.1%; Pred. No. 1.1; Matches 79; Indels 211; Gaps 34;
Matches 143; Conservative

QY 155 TDSGATWSQMT-NFPDVGTYIANPT-----DTTGYQSDIQGVVWVAFDKS--SSS 201
DB 798 TSTWASIEVSGEPESGEH---PTSGSPASGAPDTSGEPTS-----VGFELSGEQSG 848
QY 202 LQQAKTIFVGVADPNPNVFWSRDGGATWQAVPGAPTGFIPHKGVFDPVNVHLYIAT--- 258
DB 849 IGESG---LPSVDLQSSGFVPCESGLPSGDSVGLPSGIVDISGLPSAEETVSVSRIP 904
QY 259 -----SNTGPPDGGSSGDWKEFSVTSGIWTIRISVPSPSTDANDYGYSLGTTIDRQHP 310
DB 905 EVSGMPGSAESSGLHSGFSGEI-----SGT-ELISGLPSGESGLASGFP--TISLVD 955
QY 311 NTIMVAT-----QISWNPDTIIFRSTDCGATWTRIMDWTSYPNRLS 351

```

```

Db 956 TLVEVVTAAPGRQERKSGSIGVSGEELSGFPESA-----EWSSSGARG- 998
QY 352 RYVLDISAEPLMTTFGVQPNPP-----VSPKLGWMEDEAMAI DPNSRMLYGTGATL 403
Db 999 ---LPSGAE---TSGEQSGVPELSGHSQVP---GLSGEAFVPELSGHEH---SVT- 1043
QY 404 YATNDLTKWDSG-----GQIHIAPMVKGLEETA VNDL19PPSGAPLISALGDLGGFT--- 455
Db 1044 ---ELSGEHSGLPELSGEPFGVPPELSGPPSGL--DISGEFGAPEVSGPVDVSGLTSGV 1097
QY 456 ---HADVTAVPSTIFTSPVFTTGTSDYAEINPISIVR-----AGSFPPSQPNPD 502
Db 1098 DSGSEVSGV-TFISTSLQEVTTTPSVAEAEAKEILEISGLPSGETSGMVSGSLDVSGQPSG 1156
QY 503 RHVAFSTDCGKNWFOGSEPGVGT---TGTTVAASADGSRFVWAPGDPQVPVYAVGFG 557
Db 1157 -HIGF-----GGASAGVLEMSGFPFGAVESSEAS-----GVEVTSGLASG 1196
QY 558 NSWAAASQGVPAANAQIRSDRV-----NPKTFYALNSGTFRYRSTDGVTFPQPV 603
Db 1197 EESGLTSGFPTVSLVDVTTLVEVVTTQTSVAQEVGEGSGMIEISG---FLSGDRGVS--- 1249
QY 604 AAGLPSSGAVGVNTHAVPKEGDLMLAASSGLYHSTNGSSWAI TGVSSAVNVGFKSA 663
Db 1250 -----GEGSGAVQSSGLPSGTGD-FSGEPGIPYFSGDISGATDLSGQPSAVT---DISG 1300
QY 664 PGSSYPAVFVWGT--IGGVTVGAYRSDDCGTTWVLINDDQHQYGNWGOAITGDHANLRVY 721
Db 1301 EDSGLPEVTLVTSDLVEVVVTRPTVSOELG-----GETAVTPPYV 1339
QY 722 IGTVNGRGIYVGDIGGAPS 739
Db 1340 FPGSGEGSASGDLSCGAS 1357

RESULT 13
QY QUA_XANCU STANDARD; PRT; 790 AA.
AC Q9XD78;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable quinate dehydrogenase [Pyroloquinoline-quinone]
DE (EC 1.1.99.25).
GN QUMA.
OS Xanthomonas campestris (pv. juglandis).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=44291;
RN [1]_TaxID=44291;
RP SEQUENCE FROM N.A.
RC STRAIN=C5;
RX MEDLINE=20063481; PubMed=10594704;
RA Lee Y.-A., Lo Y.-C., Yu P.-P.;
RT "A gene involved in quinate metabolism is specific to one DNA homology
RT group of Xanthomonas campestris.";
RL J. Appl. Microbiol. 87:649-658(1999).
CC -I- CATALYTIC ACTIVITY: Quinate + pyrroloquinoline-quinone = 5-
CC dehydroquininate + reduced pyrroloquinoline-quinone.
CC -I- COFACTOR: PQO (By similarity).
CC -I- PATHWAY: Quinic acid catabolic pathway; first step. This pathway
CC allows growth of bacteria with quinate by its conversion to
CC protocatechuate and subsequent metabolism by the beta-ketoadipate
CC pathway.
CC -I- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -I- SIMILARITY: Belongs to the bacterial PQO dehydrogenase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC

```

entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```
CC EMBL; AF109471; AAD38453.1; -
CC InterPro; IPR001479; Bac_PQQ.
CC InterPro; IPR002372; Bac_PQQ_repeat.
CC Pfam; PF01011; PQQ; 3.
CC SMART; SM00564; PQQ; 2.
CC PROSITE; PS00363; BACTERIAL_PQQ_1; FALSE NEG.
CC PROSITE; PS00364; BACTERIAL_PQQ_2; FALSE NEG.
KW Oxidoreductase; PQQ; Quinate metabolism; Transmembrane.
FT TRANSMEM 22 42 POTENTIAL.
FT TRANSMEM 48 68 POTENTIAL.
FT TRANSMEM 77 94 POTENTIAL.
FT TRANSMEM 106 126 POTENTIAL.
SQ SEQUENCE 790 AA; 82896 MW; 87529B52A49FE6F CRC64;

Query Match 3.6%; Score 145.5; DB 1; Length 790;
Best Local Similarity 21.6%; Pred. No. 0.36;
Matches 166; Conservative 87; Mismatches 239; Indels 275; Gaps 50;

QY 60 WNNWGY--NGVVSIAADPINTNKVAAVGMVYNSWDPNDDGAILRSSDQATWQITPLPFK 117
DB 151 WSAVNTDGGSRFAALDQINRNPAG---SPGPTTPGAIANSDDGASDQLPLQ-- 205
QY 118 LGGNMPGRGGERLAVDPNNNLIYFAPSGKGLWRSSTDGATWSQMTNFPDVGTY---- 173
DB 206 -----VGEKVFCLTCHNNLALDASTGKQLWR-REINATSSVQRCRGLGVFDADA 255
QY 174 -----IANPDTTG-----YQSDIQG-VWVV-----AFKSSSSSLGQASKT 208
DB 256 ALPAPSVAAPSPIAAATVAQGANCRRLFTNTIDGLRIADADTGAFCQFGGSGNQVDLK 315
QY 209 IFVGVA-DP-----NNPVFNSRD--GGATWGAAPGAPT--GFTPHKGVFPVNVHLVIA 257
DB 316 AGLGAAPDPFYQLTSPPLVAGTIVVGGRT-RADDNVQTPMDPGVGVVSGMSVPSRS---A 370
QY 258 TSNTGGPYD-----GSS-----GDVW-KFVSITSGTWRISFV--PSTD-----TANDY 297
DB 371 GLDPGNPHDRQAAPAGSYVRSFNWAPMSYDAAMVTFLPLGGPSTDLVGAERTALDH 430
QY 298 -FGYSGLTIDRQHPNTIMVATQISWMPDTIIFRSTDDGATW-----TRIWDWTSYPNRS 350
DB 431 RYGASVLALD-----ATTGAEKWVYQTVHNDLWDF----- 460
QY 351 LRYVLDTISAEPLW-TFGVQPNPPVPSPLGWMDEMAIDPNSDRMLYGTGATLYATNDL 409
DB 461 -----DLPMQPSLIDFPNQDGSHTPAVVG----- 485
QY 410 TKWDSGGQIHIAPMVKGLEETAVNDLISPPSGA-----PLISALGDLG--GFTH 456
DB 486 TK---AGQIYVLDRTATKPLTEVREV--PVKGSIDIAHEQYAPTQPLSVGMFOIGTKHUTE 540
QY 457 ADV---TAVPST-----IFTSPTFTGTSVDYAEALNPSIIIVRAGSFDPSSQP 500
DB 541 SDMWGATAMDQMLCRIFAQKRYEGLYAP-----GTDVSL-----SF-PSGL- 582
QY 501 NDRHVAFSTDGKGNWFQSGEGVVTGT--VAASADGSRFWA---PGD----- 545
DB 583 -----GGMNW-----GGLSTDVDHVDVFPANDMRLGLWQMI PADTRKAEAAAGG 626
QY 546 -----PGQPVVAVFGNSWAASQGVPAQAQIRS--DRVNPKT---FYALNNGTF 590
DB 627 EAVNTGMGAVPLKGTPTAYV--KNRFLSALGPCQAPPYGTLSAIDLKTRSTAWQVPVGT 685
QY 591 VRSSTDGGVTFQ-PVAAGLPSSGAV-----GVMFLVAPGKGLDLAASSGLYHSTNGGSS 644
DB 686 QDTGFGIKMLPIPIGMPTLGGTILSTOGGLVF--IAGTQ-DYVLR-----FDSATGKEL 738
QY 645 WSAITGVSSAYN-VQFGKSPAGSSYPVAVVVGTTGGVTVGAYRSDDCG 690
DB 739 WKGRPLVSGSGPITVYSHKTKQY-----VVISAG---GARQSPDRG 778
```

## RESULT 14

```
Y48L_SINY3
ID Y48L_SINY3 STANDARD; PRT; 342 AA.
AC P73069;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ycf48-like protein.
GN SLR2034.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_taxid=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugiyama M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -!- SIMILARITY: Belongs to the ycf48 family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

Query Match 3.6%; Score 145; DB 1; Length 342;
Best Local Similarity 22.2%; Pred. No. 0.14;
Matches 82; Conservative 48; Mismatches 130; Indels 110; Gaps 18;

QY 350 SLRYVLDTISAEPLWTFGVQPNPPVPSPLGWMDEMAIDPNSDRMLYGTGATLYATNDL 409
DB 28 SCSHVPLDLAFNPWQBIATLDTST-----FADIAFTEDP--NHGWLWGTKETIFETD- 77
QY 410 TKWDSGGQIHIAPMVKGLEETAVNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTS 469
DB 78 -----GGDTW-----EQKLIDLGEKSAFSVFSGNEG-----ITGKPSIL--- 115
QY 470 PVFTTGTSDVYAEAL-----NPSIIIVRAGSFDPSSQPNDRHVAFSTDGKGNWFQSGEP 521
DB 116 -LHTTDDGGQTWARIPLSEKILPCAPYSIIALGQTAEMITDLGAIYKTTGGKMW-KALVE 173
QY 522 GGVTTGGTVAASADGSRFV-----WAPGD-----GQPVVAVFGNS--- 559
DB 174 GAVGVARTIQRSTDG-RYVAVSARGNFYSTWAPQGTWTPHNRNNSRRLLQTMGYGKDGQL 232
QY 560 WAASQ-----VPANAQIRSDRVNPKTFYALNNGTFYRSTDDGGVTFQPVAAAGLPSSGAV 613
DB 233 WLLARGGQLQFSTDPAEWSVDIAPO-----DKG-----SWGLL 267
QY 614 GVMFLVAPGKGLDLAASSG-LYHSTNGGSSWSAITGVSSAVNVGVFGKSPAGSSYPAVF 672
DB 268 DUSFTT-----PEEWWAGASGNLLMSQDGGQTKWAKTGV-----EDIPANLYRVVF 314
QY 673 VVGTTGGVTVG 682
DB 315 LSPKGFVLG 324


```

Search completed: May 11, 2004, 12:07:27  
Job time : 25.1745 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 11:45:55 ; Search time 86.3468 Seconds  
(without alignments)  
2704.020 Million cell updates/sec

Title: US-09-917-376-3

Perfect score: 4036

Sequence: 1 ATTPQYTSNVAIGGGFVD.....YIGTNGRGIVYDGGAPSG 740

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_25:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phage:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_rvirus:\*

16: sp\_bacteriap:\*

17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2429.5	60.2	882	16 Q82M04	Q82M04 streptomyc
2	2371	58.7	890	16 Q86727	Q86727 streptomyc
3	2062.5	51.1	996	2 Q9AQH0	Q9AQH0 caldicellul
4	2009	49.8	839	16 Q97KK0	Q97KK0 clostridium
5	1765.5	43.7	857	3 Q8TFP1	Q8TFP1 aspergillus
6	1753.5	43.4	806	3 Q9P4T8	Q9P4T8 agaricus bi
7	1680	41.6	856	3 Q74170	Q74170 aspergillus
8	1579	39.1	838	3 Q7Z9M8	Q7Z9M8 trichoderma
9	1405	34.8	739	16 Q82K30	Q82K30 streptomyc
10	1162.5	28.8	751	16 Q8PLM5	Q8PLM5 xanthomonas
11	1160.5	28.8	707	16 Q9WE1	Q9WE1 xanthomoga
12	1145	28.4	751	16 Q8P9U5	Q8P9U5 aspergillus
13	1117	27.7	812	3 Q8J0D2	Q8J0D2 geotrichum
14	409	10.1	196	3 Q8J1H7	Q8J1H7 agaricus bi
15	368.5	9.1	1064	2 Q7X2U2	Q7X2U2 uncultured
16	238	5.9	2468	16 Q9I2M3	Q9I2M3 pseudomonas

17	214	5.3	1800	2	Q9L948	Q9L948 pseudomonas
18	214	5.3	8682	16	Q88RG2	Q88RG2 pseudomonas
19	205	5.1	2215	16	Q7WBN0	Q7WBN0 bordetella
20	202.5	5.0	2174	16	Q92UU8	Q92UU8 rhizobium m
21	198.5	4.9	3346	16	Q7WNS4	Q7WNS4 bordetella
22	197	4.9	2016	17	Q8TJ59	Q8TJ59 methanosarc
23	189	4.7	3972	16	P73139	P73139 synechocyst
24	186.5	4.6	4199	16	P74440	P74440 synechocyst
25	185.5	4.6	818	6	Q9N1P0	Q9N1P0 bos taurus
26	185.5	4.6	997	2	Q9Z411	Q9Z411 bacillus sp
27	184.5	4.6	998	17	Q9HL09	Q9HL09 thermoplasma
28	184	4.6	805	15	Q88GZ6	Q88GZ6 pseudomonas
29	183	4.5	2523	17	Q8TJ58	Q8TJ58 methanosarc
30	182.5	4.5	848	16	Q9RK65	Q9RK65 streptomyc
31	182	4.5	1070	16	Q8XW00	Q8XW00 ralstonia s
32	182	4.5	1904	9	Q38319	Q38319 lactococcus
33	179	4.4	2079	16	Q7UF08	Q7UF08 rhodopirell
34	178.5	4.4	3420	16	Q8FUS1	Q8FUS1 brucella su
35	177	4.4	666	11	Q8BRF6	Q8BRF6 mus musculu
36	177	4.4	928	11	Q8OUN3	Q8OUN3 mus musculu
37	177	4.4	1462	11	Q80WR0	Q80WR0 mus musculu
38	175.5	4.3	2523	16	O53393	O53393 mycobacteri
39	174.5	4.3	3930	16	Q98E20	Q98E20 rhizobium l
40	174	4.3	1271	17	Q8T155	Q8T155 methanosarc
41	172	4.3	993	16	Q8YYQ5	Q8YYQ5 anabaena sp
42	171.5	4.2	344	16	Q7W1M1	Q7W1M1 bordetella
43	171.5	4.2	7716	16	Q7UWZ8	Q7UWZ8 rhodopirell
44	171	4.2	908	16	Q8YR62	Q8YR62 anabaena sp
45	170.5	4.2	344	16	Q7WPL2	Q7WPL2 bordetella

## ALIGNMENTS

### RESULT 1

Q82M04	ID	Q82M04	PRELIMINARY;	PRT;	882 AA.
AC	Q82M04;				
DT	01-JUN-2003	(TREMREL. 24, Created)			
DT	01-JUN-2003	(TREMREL. 24, Last sequence update)			
DT	01-OCT-2003	(TREMREL. 25, Last annotation update)			
DE	Putative endo-1,4-beta-glucanase.				
GN	CELA3 OR SAV1856.				
OS	Streptomyces avermitilis.				
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;				
OC	Streptomycineae; Streptomycetaceae; Streptomyces.				
OX	NCBI_TaxID=33903;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;				
RX	MEDLINE=21477403; PubMed=11572948;				
RA	Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,				
RA	Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,				
RA	Kikuchi H., Shiba T., Sakaki Y., Hattori M.;				
RT	"Genome sequence of an industrial microorganism Streptomyces				
RT	avermitilis: deducing the ability of producing secondary				
RT	metabolites.";				
RL	Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;				
RX	MEDLINE=22608306; PubMed=12692562;				
RA	Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,				
RA	Sakaki Y., Hattori M., Omura S.;				
RT	"Complete genome sequence and comparative analysis of the industrial				
RT	microorganism Streptomyces avermitilis.";				
RL	Nat. Biotechnol. 21:526-531(2003).				
DR	EMBL; AP005028; BAC69567.1;				
DR	GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.				
DR	GO; GO:0005975; P:carbohydrate metabolism; IEA.				
DR	InterPro; IPR001919; Bac celose-bind.				
DR	InterPro; IPR008965; Cellul bind.				
DR	InterPro; IPR002860; GH_BNR.				

DR Pfam; PF02012; BNR; 10.  
DR Pfam; PF00553; CBM\_2; 1.  
DR SMART; SM00637; CBD\_II; 1.  
KW Complete proteome.  
SQ SEQUENCE 882 AA; 92152 MW; A3C8E753CE69F13D CRC64;

Query Match 60.2%; Score 2429.5; DB 16; Length 882;  
Best Local Similarity 59.0%; Pred. No. 1.6e-126;  
Matches 438; Conservative 111; Mismatches 174; Indels 19; Gaps 8;

QY 6 YTSNVAIGGGGFGVIGNECAPGILYVTRDIGMYRWDAAANGRWIPILLDWGNNWGY 65  
Db YSKNARVDGGGFGVIGNESEKRLAYARTDIGAYNEABSSKTTWTLDSVGVGDWGH 98  
QY 66 NGVSVIAADPINTNKVAAVGMVNTSNPDNDGAILRSSDQAGTWOITPLPFKLGNNMGR 125  
Db TGVVSLASDVDPNKVYAAVGTNTNSWDPGNGAVLRSGDRGASQKTDLPFKLGNNMGR 158  
QY 126 GNGERLAVDPNNNTLYFGAPSGKGLWRSTDSGATWSQMTNPPDPVGTIYANPTDTTGYQS 185  
Db 159 GNGERLAVDPNRSVLYLGAPSGKGLWRSTDSGASWSQVTFPNVGTIVQDADTDSGYAS 218  
QY 186 DIQGVVWVAFDKSSSLGQASKTIFVGVADPNPNVFNWSDRGATWQAVPGAPTGFPHKG 245  
Db 219 DNQGIWVTFDESTGSGSTRRTVYGVADKDNSVYRSTDRAGATWSRLAGQPTGHLAHK 278  
QY 246 VFDPNVHLYIATNTGPGYDSSGDVWKFVSITSGTWTRISGVPSTDTANDYFGYSGTLTI 305  
Db 279 VLDAANGCLYLAYSQKPGYDGGKQLWRYTKTGTNTNISFVABADT---YFGSGLTIV 335  
QY 306 DRQHPNTIMATQISWPPDTTIFRSTDCGATWTRIDWTSYPNRSLRYVLDISAEPWLTFF 365  
Db 336 DRQHPGTWATAYSWPPDTQLFRTSDSGGTWKAMDVTSYPSRSNRFTMDVSSSEPLTW 395  
QY 366 GVQNPVPVPSKLGWMDAMADPNRSDMLYGCATLYATNDLTKWDSGGQIHTAPWVK 425  
Db 396 GANPAPPQTPKLGWMTSLSLIDPPDSARMYGTGATVYGTNDLNTWDSGSGQTIKPMAR 455  
QY 426 GLEETAVNDLISPPS-GAPLISALGDGFFTHADVTAVPSTFTSPVTTGTSVDYAEIN 484  
Db 456 GLEETAVNDLASPPSGGQLFSLGDIGGFRHTDUTVPSLMYTSFNTTSTSLDYAETD 515  
QY 485 PSIIIVRAGSFPSSQPNDRHVAFTSDGGKWNFGQSEPGVTTGGTVAASADGRFVWAPG 544  
Db 516 PGTVVRVGNLD--SGP---HVAFTDNGANFAGADPSVSGGTVAASDGRFVWSPA 570  
QY 545 DPGQPVTVAVFGNSWAASQGVPAQAQTRSRVNPKTFFYALNSGTFFYRSTDCGVTFPQVA 604  
Db 571 GTG--VOYTTGFTGTSWASAGLPAGAI VESDRVDPKTFYGFKSGRFYVSSDGGATFTASA 628  
QY 605 A-GLPSSGAVGMFHAVPGKGDWLAASS-----GLVHSTNGGSSWSAITGVSSAVNVG 658  
Db 629 ATGLPSSGDS--VRFKALPGTGDILWLAGASDGAAYGLWHSTDGGAAFTKLAIVDQADTIG 686  
QY 659 FGKAPGSSYPVAVFVVGITGGVTAIRSDDCCTTWLINDQHQHVGNNQCAITGPHANLR 718  
Db 687 FGKAATGASYQTLTYSAKIGVGRGIFRSTDKASWTRVNDAAHQWGTGAALITGDPVRYG 746  
QY 719 RVYGTNGRGIVYGDIGAPSG 740  
Db 747 RVYVSTNGRGIVYGDTAGSSDG 768

RESULT 2  
O86727 ID  
AC O86727; PRELIMINARY; PRT; 890 AA.  
DT 01-NOV-1998 (Tremblrel. 08, Created)  
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)  
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
DE Putative secreted cellulase.  
GN SC06545 OR SC5C7.30C.  
OS Streptomyces coelicolor.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Streptomycineae; Streptomycetaceae; Streptomyces.  
RN [1] NCBI\_TaxID=1902;  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Seeger K.J., Harris D.;  
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Parkhill J., Barrell B.G., Rajandream M.A.;  
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RX MEDLINE=97000351; PubMed=8843436;  
RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,  
RA Kinashi H., Hopwood D.A.;  
RT "A set of ordered cosmids and a detailed genetic and physical map for  
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
RL Mol. Microbiol. 21:77-96(1996).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2) / M145;  
RX MEDLINE=21996410; PubMed=12000953;  
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,  
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
RA Rabinowitch E., Rajandream M.A., Rutherford K., Rutter S.,  
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,  
RA Hopwood D.A.;  
RT "Complete genome sequence of the model actinomycete Streptomyces  
coelicolor A3(2).";  
RL Nature 417:141-147(2002).  
DR EMBL; AL939128; CAA20642.1; --  
DR PIR; T35237; T35237.  
DR HSP; P07986; IEKQ.  
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.  
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
DR InterPro; IPR001919; Bac celose-bind.  
DR InterPro; IPR008965; Cellul bind.  
DR InterPro; IPR02860; CH\_BNR.  
DR Pfam; PF02012; BNR; 10.  
DR Pfam; PF00553; CBM\_2; 1.  
DR SMART; SM00637; CBD\_II; 1.  
KW Complete proteome.  
SQ SEQUENCE 890 AA; 93252 MW; D2C58695A4856E84 CRC64;

Query Match 58.7%; Score 2371; DB 16; Length 890;  
Best Local Similarity 57.7%; Pred. No. 2.8e-123;  
Matches 431; Conservative 109; Mismatches 183; Indels 24; Gaps 10;

QY 6 YTSNVAIGGGGFGVIGNECAPGILYVTRDIGMYRWDAAANGRWIPILLDWGNNWGY 65  
Db 41 YTKNARVDGGGFGVIGNESEKRLAYARTDIGAYNEABSSKTTWTLDSVGVGDWGH 100  
QY 66 NGVSVIAADPINTNKVAAVGMVNTSNPDNDGAILRSSDQAGTWOITPLPFKLGNNMGR 125  
Db TGVVSLASDVDPNKVYAAVGTNTNSWDPGNGAVLRSGDRGASQKTDLPFKLGNNMGR 160  
QY 126 GNGERLAVDPNNNTLYFGAPSGKGLWRSTDSGATWSQMTNPPDPVGTIYANPTDTTGYQS 185  
Db 161 GNGERLAVDPNHDVLYLGAPSGHGLWRSTDSGATWSQMTNPPDPVGTIYANPTDTTGYQS 220  
QY 186 DIQGVVWVAFDKSS--SSLGQASKTIFVGVADPNPNVFNWSDRGATWQAVPGAPTGFPHK 244  
Db 221 DNQGIWVTFDESTGSGAGTATRTLYGVADKENAVYRSTDRAGATWERLAGQPTGYLAHK 280  
QY 245 GVFDPPNVHLYIATNTGPGYDSSGDVWKFVSITSGTWTRISVVPSTDTANDYFGYSGLT 304

```
Db 281 GVLDAENGLYLAYSDDGGPVDGKGLRYATATGTWTDISPAAEDT---YGFSGLT 337
Qy 305 IDRHQNTIMVATOISWMPDITIIERSDGGATWTRIMDWTSPNRSIRYVLDISAEPWLT 364
Db 338 VDRQPGTWMATAYSSWMPDITIIERSDGGATWTRIMDWTSPNRSIRYVLDISAEPWLT 397
Qy 365 FGVQNPVPSPKLGWMDMAIDPFNSDRMLYGTGATLYATNDLTKW-DSGGQIHIAPM 423
Db 398 WGANPAPPEQTPKLGWMTALEIDPFDSDRMYGTGATVYGTENLTNWDDGGTFAVEPM 457
Qy 424 VKGLEETAVNDLISPPSGAPLISALGDLGGTHADVTAVPSTIPTSVPFTTGTSDVDAEL 483
Db 458 VRGLEETAVNDLISPPSGAPLISALGDLGGTHADVTAVPSTIPTSVPFTTGTSDVDAEL 517
Qy 484 NPSIIVRAGSDPPSQPNDRHVAFTDGGKWNFGSEBPGVYTGCTVAASADGSRFWWAP 543
Db 518 KPDVVVRAGNDJ--SGP---HIAFTDNGANWFGDTPSGVGGTVAAGADGSRFWWSP 572
Qy 544 GDPQPVVYAVFGNSMAASQGVPAQAQIRSDRVNPKTFYALSNGTFYRSDGGVTFQPV 603
Db 573 --EGAGVQYTTGFGTSMQASTGLPAGALVESDRVNPATFYGKSGRFVVDGGATFTAS 630
Qy 604 AA-GLPSSGAVGVMFHVPKGLWLAASS-----GLYHSTNGSSWSAITGVSSAVNV 657
Db 631 AATGLPAGD--GVRFKALPGEGDVWLAGGAADGPYGLWHSTDGGGTFTRLPGVDAADTV 688
Qy 658 GFGKSAPGSSYPVAVVGTIGVTVGYSDDCGTWWLINDDOHQYGNWGAITGDHANL 717
Db 689 GFGKAAPGASQTLFTSABIGGVNGIFRSTDAGATWTRVNDADAHQWGTGAATIGDPRVY 748
Qy 718 RRVVIGTNGRGIVYVDI-----GGAPSG 740
Db 749 GRVYVATNGRVIYVDISDTGGGDPG 775
```

## RESULT 3

```
Q9AQHO PRELIMINARY; PRT; 996 AA.
AC Q9AQHO;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Glycosyl hydrolase 5 (fragment).
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tok7B.1;
RX MEDLINE=20171169; PubMed=10706665;
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
RT thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL Curr. Microbiol. 40:333-340(2000).
DR EMBL; AF078038; AAK06388.1; -.
DR HSSP; Q06851; INEC.
DR GO; GO:0016787; P:hydrolase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR002860; GH_BNR.
DR Pfam; PF00942; CBM_3; 1.
DR ProDom; PD001947; CBD_3; 1.
KW Hydrolase.
FT NON_TER 996
SQ SEQUENCE 996 AA; 108275 MW; 3C72B6BD22F3C614 CRC64;
```

## Query Match

Best Local Similarity 51.1%; Score 2062.5; DB 2; Length 996;

Pred. No. 4.1e-106;

```
Matches 387; Conservative 96; Mismatches 238; Indels 15; Gaps 9;
Qy 4 QPYTWSNVAI-GGGFVDGIVFNEGAPGLIYVTRDIGMYRWDAAANGRWIPLLDWYGNW 62
Db 35 QPYVKNVKIEGGGFTIGIVFNPKKNLVYVTRDIGAYRSTGGNTWTQLMDWVSFDD 94
Qy 63 WYNGVNSIAADPINTKV-WAANGMYTNSWDPNDGAILRSSDQCATWITPLPFKLGN 121
Db 95 WNLGVESIATDPVDPNVILACQGYTNSWTDMMGAILRSTDEGDTFFITPLPFKLGN 154
Qy 122 MPGRGMGERLAVDPNNNDILYFGAPSGKGLWRSTDSGATWSQMTNPPDVCTYIANPTDTT 181
Db 155 MPXRNIGERLAIIDPNNRILYLGREGKGLWSEBVGWKKVTSFPNPGTYIEDNCPN 214
Qy 182 GYQSDIQGVVWVAFDKSSSSLGQASKTIFVGVADPNNPVFVRDGGATWQAVPGATGFI 241
Db 215 DYLNIHTGVWVVDPTSGRPGEGXKIIYGVADXTTSIYYTKDGGQTWQALPGQPTGLL 274
Qy 242 PHKGVDPVNVHLYIATNTGGPYDGGSDWVKFVSTGTSWTRISVPSTDTTANDYFGYS 301
Db 275 FORAKLSS-DGMLYITYTSNTQGYNDYGEVWRYNTKTGEWKNISPMQAQDT---YFGY 330
Qy 302 GLTIDROHPNTIMVATOISWMPDITIIFRSTDCGATWTRIMDWTSPNRSIRYVLDISAEP 361
Db 331 GLAVDAQNPVKVWVAALSSWMPDITYIWRSIDGGETWKCIEWWNGYFNRTLHYNNDISAEP 390
Qy 362 WLTFG-VQPNPVPSPKLGWMDMAIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420
Db 391 WLNFXETBTPPEVSPKLGWVGTLEIDPFNSDXMLYGTGAXLYGCCDLLTNWDKQNTI 450
Qy 421 APMVKGLEETA VNDLISPPSGAPLISALGDLGGTHADVTAVPSTIFTSPFTTGTSDVY 480
Db 451 KYKAI GIEETSVOALISPPVGPPLFSALGDIAGFRHEDLEKAPNWTYVQPNMGTITDIF 510
Qy 481 AELNPSIIVRAGSDPPSQPNDRHVAFTDGGKWNFG-SEPGVITGCTVAASADGSRF 539
Db 511 AELNPSIIVRAGSDPPSQPNDRHVAFTDGGKWNFG-SEPGVITGCTVAASADGSRF 570
Qy 540 VWAPGDPQPVVYAVFGNSMAASQGVPAQAQIRSDRVNPKTFYALSNGTFYRSDGGVT 599
Db 571 VWAP--KGAKVCYTDNGKNVCEANVPSEALIVSDRVNPKTFYAFNGKFYISADGKT 628
Qy 600 F-QPVAAGLPSSGAVGVMFHVPKGLWLAASSGLYHSTNGSSWSAITGVSSAVNVG 658
Db 629 FIESPAAGLPISG----NFKTVPGIEGDIWLVGNNGMWHSTDGGYSFVKISGVDEAASIG 684
Qy 659 FGSAPGSSYPVAVVGTIGVTVGYSDDCGTWWLINDDOHQYGNWGAITGDHANLR 718
Db 685 FGSAPGSSYPVAVVGTIGVTVGYSDDCGTWWLINDDOHQYGNWGAITGDHANLR 744
Qy 719 RYVIGTNGRGIVYVDI 734
Db 745 RVFVATNGLGIKWGEI 760
RESULT 4
Q97KKO PRELIMINARY; PRT; 839 AA.
AC Q97KKO;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Probably secreted sialidase, several ASP-boxes and dockerin
DE domain.
DE CAC0919.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
```



RA Noelling J., Breton G., Omselchenko M.V., Makarova K.S., Zeng Q.,  
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,  
RA Tatusov R.L., Sabatne F., Doucette-Stamm L., Soucaille P., Daly M.J.,  
RA Bennett G.N., Koonin E.V., Smith D.R.;  
RT "Genome sequence and comparative analysis of the solvent-producing  
bacterium Clostridium acetobutylicum";  
RL J. Bacteriol. 183:4823-4838(2001).  
DR EMBL; AE007608; AAK78895.1; -.  
DR F01; D97013; D97013.  
DR GO; GO:0005509; F:calcium ion binding; IEA.  
DR GO; GO:0000272; P:polysaccharide catabolism; IEA.  
DR InterPro; IPR002105; Dockerin\_1.  
DR InterPro; IPR002048; EF-hand.  
DR InterPro; IPR002860; GH\_BNR.  
DR Pfam; PF02012; BNR; 9.  
DR Pfam; PF00404; Dockerin\_1; 2.  
DR PROSITE; PS00018; EF\_HAND; 2.  
KW Complete proteome.  
SQ SEQUENCE 839 AA; 90824 MW; 6C99A041CA9CF984 CRC64;  
  
Query Match 49.8%; Score 2009; DB 16; Length 839;  
Best Local Similarity 50.9%; Pred. No. 3e-103;  
Matches 377; Conservative 107; Mismatches 240; Indels 16; Gaps 12;  
  
QY 3 TQPYTWSNVAIGGGFVDGIVNEGAPGILYVRTDIGMYRWDAANGRWIPLLDWGNN 62  
DB 37 SQGYKWDNAKIGAGYVPAVFNKTEKOLYARTDMGAYRWDAANGRWIPLLDWGNN 94  
QY 63 WYNGVVSIAADPTINTKVAAGMYTNSWDNDGAILRSSDQATWQITPLPKLGGNN 122  
DB 95 WTMJGCSIAIDPTDITNRVYIAAGLYTNDQENAYILSSQDKGNTWKRYQPLPKVGGNN 154  
QY 123 PGRGGERLAVDPNNNDILYFGAPSGKGLWRSTDSGATWSQMTNPPDVGTYIANPTDITG 182  
DB 155 PGRNGERLQIDPNDKILYLARGSGNLKSEYDGTQSKVDNPDYDGYVDQNE-- 212  
QY 183 YQSDIQGVVWVAFKSSSLGQASKTIIFGVVAD-PNNPVFWSRDCGATQVAVPGAPTGI 241  
DB 213 YTADKGVGVVWETFDPTSTGKSTPTQMTYVGAADTKGNNIYVNDGKTSWASVKGP 272  
QY 242 PHKGVPDPVNHVLYATNTGPGYDSSGDVWKFVSTGTTWTRISPVSTDTANDYFGVS 301  
DB 273 PHHGLL-ASDGMWLYISYNTGPGYDSSGDVWKFVSTGTTWTRISPVSTDTANDYFGVS 328  
QY 302 GLTIDRQHPNTIMVATQISWPDITIFRSTDCGATWTRIDWTSYPNRSLRVLDISAEP 361  
DB 329 GISVDAQPNVAVATLNRWPFDEIYRSTDAKTWPKPIWNGVYPNRTLGNLDYSAGP 388  
QY 362 WLTFG-VQNPFPVSPKLGWDEAMAIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420  
DB 389 WLDWKTGVTTPDPLVPLKLGWMMGLDIDPFNSDRMFYGTGATLYGTDDLTKWDKGNVDI 448  
QY 421 APVKGLLEETAVNDLISPPSGAPLISALDGLGFFHADVTAVPSTIFTSPTTGTSDY 480  
DB 449 SVKANGIEECVANDVVPVTKAQLLSAVDGCDFYHDDITKVPKQMTNPFSAITSDY 508  
QY 481 AELNPSIIVRAGSDPSSQPNDRHVAFTDGGKWNFO-GSEPGVYTCGTAAASADGSRF 539  
DB 509 AESVNFVVRVGVNDTSKNQDQDCGSIYDGGKWNFSAGNISGVYKAGTVAAGADAKTI 568  
QY 540 VWAPDGPQPVVYAVGFGNSWAASQGVPAQAIRSDRVNPKTFYALSNGTFYRSTDGGVT 599  
DB 569 VWSP-EEGANAAYSTDNGKNTPCSLGPGQAKVRSRDRVNPCKFYFLNGKFIISTDAGAT 627  
QY 600 F-QPVAAGLPSGAVGVWFAVPGKEGLWLA-ASSGLYHSTNGSSWSAITGVSSAVNV 657  
DB 628 FTQSSQTGLPTKGR-GI-FKTVIGHEGDIWTAGGKDLWHSTDSGATFTKVSQVSDASDTV 685  
QY 658 GFQKSAPGSSYPAVFVCTIGVTVGAYSDCGTTWVLINDDHOQYGNWQAIITGDHANL 717  
DB 686 GLGSKTDGDPALYMDATIDGTAGIFRSDDEGATWVRINDDAHQYSGPDYCIITGDPNKKY 745  
QY 718 RRVVIGTNGRGIVYGDIGGA 737

DB 746 GRFVGTNGRGIVYGDIGGS 765  
RESULT 5  
Q8TFPI PRELIMINARY; PRT; 857 AA.  
AC Q8TFPI; 857 AA.  
DT 01-JUN-2002 (TREMELrel. 21, Created)  
DT 01-JUN-2002 (TREMELrel. 21, Last sequence update)  
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)  
DE Endoglucanase C.  
GN EGLC.  
OS Aspergillus niger.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
OC NCBI\_TaxID=5061;  
RN [1]\_TaxID=5061;  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21914059; PubMed=11916668;  
RA Hasper A.A., Dekkers E., van Mil M., van de Vondervoort P.J.I.,  
RA de Graaff L.H.;  
RT "EgIC, a New Endoglucanase from Aspergillus niger with Major Activity  
towards Xyloglucan";  
RL Appl. Environ. Microbiol. 68:1556-1560(2002).  
DR EMBL; AY040839; AAK77227.1; -.  
DR GO; GO:0005576; C:extracellular; IEA.  
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.  
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
DR InterPro; IPR000254; CBD\_fungal.  
DR InterPro; IPR002860; GH\_BNR.  
DR Pfam; PF02012; BNR; 7.  
DR ProDom; PD001821; CBD\_fungal; 1.  
DR SMART; SM00236; fCBD; 1.  
DR PROSITE; PS00562; CBD\_FUNGAL; 1.  
SQ SEQUENCE 857 AA; 90459 MW; 482253BCD919C511 CRC64;  
  
Query Match 43.7%; Score 1765.5; DB 3; Length 857;  
Best Local Similarity 47.1%; Pred. No. 9.6e-90;  
Matches 351; Conservative 107; Mismatches 258; Indels 29; Gaps 14;  
  
QY 1 ATTQPYTWSNVAI-GGGFVDGIVNEGAPGILYVRTDIGMYRWDAANGRWIPLLDWG 59  
DB 20 AASQAYTKNVVTVGGGGTTPGIVFNPSAKGVAYARTDIGAYRLN-SDDTWTPLMDWAN 78  
QY 60 WNNWYGVVSTAADPTINTKVAAGMYTNSWDNDGAILRSSDQATWQITPLPKLG 119  
DB 79 NSNMHDWDGIDALATDPDVTDRVYVAVGMYTNDWDNDGAILRSSDQATWQITPLPKLG 138  
QY 120 GMPGRCGERLAVDPNNNDILYFGAPSGKGLWRSTDSGATWSQMTNPPDVGTYIANPTD 179  
DB 139 GMPGRCGERLAVDPNNNDILYFGAPSGKGLWRSTDSGATWSQMTNPPDVGTYIANPTD 198  
QY 180 TTGYOSDIOGVVWVAFKSSSLGQASKTIIFGVVADPNNPVFWSRDCGATQVAVPGAPT- 238  
DB 199 T--YTSDPVGIWVAFDSTSGSGSPTPIIFGVVADPNNPVFWSRDCGATQVAVPGAPT- 256  
QY 239 GPIPHKGVDPVNHVLYATNTGPGYDSSGDVWKFVSTGTTWTRISPVSTDTANDYF 298  
DB 257 GLPLHKGLSPSEHTLIYISYNGAGPYDGTNGTVHKYNTSGVMTDISPTSMTDT---YY 313  
QY 299 GYSGLTIDRQHPNTIMVATQISWPDITIFRSTDCGATWTRIDWTSYPNRSLRVLDIS 358  
DB 314 GYGGGLAVDQVPGTVNVAALNCWMPDELIRWSTDSGATWSQMTNPPDVGTYIANPTD 373  
QY 359 ABFWLITFGVQPNP-PVPSPLKGMDEAMAIDPFNSDRMLYGTGATLYATNDLTKWDSGGQ 417  
DB 374 NAFWLQDSTDSDFPV--RVGMVWEALAIIDPDSHNLVGTGETIYGGHDLQNWDSHN 430  
QY 418 IHAPMVKLEETAVNDLISPPSGAPLISALDGLGFFHADVTAVPSTIFTSPTTGTGTS 477  
DB 431 VTIESLAVGIEEMAVLGLITPPGPGALLSAVGDGCGFYHTSLTAPSQYHTPTYSSTNG 490

```

Qy 478 VDYAELNPSIIVRAGSDPPSQPNDRHVAFSTDCGKNWFQSSBPGVTTGGTVAASADGS 537
Db :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
491 IDYAGNKPANIVRSGSSD-----SDPTIALSSSFGESWYAAASSSTATQVALSADAD 545
Qy 538 RFVWAPGDPGPVVYAVGFGNSWAASQGVPANAOIRSDRVNPKTFYALNSGTFTYRSTDDG 597
Db :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
546 TILMNSDG-----AYRSANSATLSAVSLPSGAVIASDKANTTFYAGSSSFYLSDDTA 601
Qy 598 VTFOPVAAGLPSSGAVGMFHAVPGKEGDLWLAASSGLYHSTNGSSWSAI-TGVSSAVN 656
Db :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
602 ATP-TVTTILGSSSTANAI-PAQPSLAGDVWVSTDTGLFHSNTYKGFQIGSCTEGWS 659
Qy 657 VGFKSAPGSSYPVAVVGTIGTGYAYRSDDCGTTWVLNDDQHOYCN-WGQAITGDHA 715
Db :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
660 FGFKPSSDGYPVLFPGFTVDGTLGFKTEQGVNWQIISDAEHGFGSASANVVGDLQ 719
Qy 716 NLRVYIGTNGRGIVYGDIGAPSG 740
Db :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
720 NYGRVFGTNGRGIFYGD-----PSG 740

RESULT 6
Q9P4T8 PRELIMINARY; PRT; 806 AA.
AC Q9P4T8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE CEL6 protein precursor.
GN CEL6.
OS Agaricus bisporus (Common mushroom).
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Agaricales; Agaricaceae; Agaricus.
OX NCBI_TaxID=5341;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D649;
RA Morales-Almora P., Thurston C.F.;
RT "Molecular analysis of the cellulolytic genes in Agaricus bisporus.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ292929; CAC02964.1; -.
DR HSSP; P00725; 2CBH.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0004553; F:hydrolyase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR00254; CBD_fungal.
DR Pfam; PF02012; BNR; 8.
DR SMART; SM00236; fCBD; 1.
DR PROSITE; PS00562; CBD_FUNGAL; 1.
KW Signal.
FT CHAIN 1 23 POTENTIAL.
FT CHAIN 24 806 CEL6.
SQ SEQUENCE 806 AA; 84779 MW; 7C67B382B62FF341 CRC64;

Query Match 43.4%; Score 1753.5; DB 3; Length 806;
Best Local Similarity 45.9%; Pred. No. 4.1e-89;
Matches 341; Conservative 112; Mismatches 249; Indels 41; Gaps 10;

Qy 1 ATTQPTWSNVAI-CGGGFVDGIVNEGAPGILYVRTDIGMYRWDANGRWIPLLDVGW 59
Db :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
23 AASQYRWQNVKYGGGGFGVGFVFNSEKGLAYARTDIGAYKLN-ADDTWTPLLDPAD 81
Qy 60 WNNNGYGVWSIAADPINTNKVMAAVGMYTNSWPNDCGAILRSSDQAGTQWITLPPFKLG 119
Db :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
82 DSRWNYGVVDALADVPQPNRLYLATGLYTNWSPNPSNGHILISDYGKTFDAAPLPFKLG 141
Qy 120 GNMFGRCMGERLAYDPNNNDNTLYFCAPSCKGLWRSTDSGTWQVTFPDVGTIYANPTD 179
Db :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
142 GNMFGRCMGERLVVDPNPLNSILYFCARSGNGLWKSSTNSGRWSKVTSFTDTGSPVDPSPD 201

```

```

Qy 180 TTGYQSDIQGVVWVAFDKSSSSLSQASKTIIFGVADP-NNPVFWSRDGGATWQAVPGAPT 238
Db :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
202 STLNSDKIGIAVWITIDKASGSSGATPRIFGVANKGNSNIYISTNGSSWSAVAGQPT 261
Qy 239 GFIPHKGVDFVNVHLYIATSNITGPGYDGGSDGVWKFVSTGTTWTRISVPVSTDTANDYF 298
Db :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
262 SFLPHKGVLSERALYITYSDCGGPGYDGTSAVYKISIDTGAWTDITPVSGSDL---FF 318
Qy 299 GYSGLTIDRQHPNTIMVATQISWPDITIFRSTDCGATWTRIDWMTSPNRSURYLDIS 358
Db :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
319 GFGGLADTQKNGTVMVAALNSWPDGQIFRSTNGGASWTLMDMAVYPLTNKYYSNAA 378
Qy 359 AEPMLTTFGVQVQNPVPSP---KLGMWDEMAIDPPNSDRMLYGTGATLYATNDLTWDSG 415
Db :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
379 LAPW-----IGENIVDVTPCNLIQIWMMSLSIDPFDNHLWLYGTGETIYCSRLLLKWSA 434
Qy 416 GQIHAPVKGLEFAVNDLISPPSCAPLISALGDLGGETHADVTAVPSTIFTSVPVFTTG 475
Db :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
435 HNVTKSLADGVEETSVOALLISPPSGPLVSAIGDVGVGFHOSLTRAPSAQFTNPKWSTS 494
Qy 476 TSDVYAEINPSIIVRAGSFPDSSQPNDRHVAFSTDCGKNWFQSSBPGVTTGGTVAASAD 535
Db :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
495 ADIDFAGNPNPNVRIGTGDSTS---GKQVAISSDYGVTVNQHFGAPDNVQGGKVAISAD 551
Qy 536 GSRFVWAFGDPGPVVYAVGFGNSWAASQGVPANAOIRSDRVNPKTFYALNSGTFTYRST 595
Db :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
552 ADIILWRTNGNG-----VWVSRNQATFNIFYGASGKTFYVSTD 589
Qy 596 GGVTFQPVAAGLPSSGAVGMFHAVPGKEGDLWLAASSGLYHSTNGGSSWSAITGVSSAV 655
Db :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
590 NGKTFSAHSGSLGSATSVFDITVH--PSVSGDIWASTDKGLFHTSDSGATFSAISGTOAW 647
Qy 656 NVGFKSAPGSSYPVAVVGTIGTGYAYRSDDCGTTWVLNDDQHOYCN-WGQAITGDH 714
Db :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
648 GVALGAPRSTGCGYPAVFAAANYGGEVAYFRSDRGVNVWVKINDAHGFGAASANCRAADP 707
Qy 715 ANLRVYIGTNGRGIVYGDIGGA 737
Db :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
708 RVYGRVYIGTNGRGIFYGDVAGS 730

RESULT 7
Q74170 PRELIMINARY; PRT; 856 AA.
AC Q74170;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Avicelase III.
GN AVIII.
OS Aspergillus aculeatus.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5053;
RN [1]
RP SEQUENCE FROM N.A.
RA Arai M., Takada G., Kawaguchi T., Sumitani J.;
RT "Avicelase III from Aspergillus aculeatus.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB015511; BAA29031.1; -.
DR PIR; T00349; T00349.
DR HSSP; P00725; 2CBH.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0004553; F:hydrolyase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR00254; CBD_fungal.
DR InterPro; IPR002860; GH_BNR.
DR Pfam; PF02012; BNR; 7.
DR Pfam; PF00734; CBM_1; 1.
DR ProDom; PD001821; CBD_fungal; 1.
DR SMART; SM00236; fCBD; 1.
DR PROSITE; PS00562; CBD_FUNGAL; 1.
SQ SEQUENCE 856 AA; 89820 MW; BE085983AF60ED76 CRC64;

```



```
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
DE 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DB Putative glycosyl hydrolase.
GN SAV2574.
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948;
RA Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis: deducing the ability of producing secondary
RT metabolites.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=22608306; PubMed=12692562;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.;
RT "Complete genome sequence and comparative analysis of the industrial
RT microorganism Streptomyces avermitilis.";
RL Nat. Biotechnol. 21:526-531(2003).
DR EMBL; AP005031; BAC70285.1; -
DR GO; GO:0016787; F:Hydrolase activity; IEA.
DR InterPro; IPR002860; GH_BNR.
DR DR InterPro; IPR006311; Tat.
DR Pfam; PF02012; BNR; 9
DR TIGRFAMs; TIGR01409; Tat_signal_seq; 1.
DR Hydrolase; Complete proteome.
KW SEQUENCE 739 AA; 78399 MW; 8E3B0F3D0F5B3D28 CRC64;

Query Match 34.8%; Score 1405; DB 16; Length 739;
Best Local Similarity 40.8%; Pred. No. 7.5e-70; Indels 86; Gaps 21;
Matches 308; Conservative 104; Mismatches 257;

QY 6 YTSNVAIGGGFVDGIVFNEGAGFGLYVRDIDGMYRWDAANGRWIPLLDWGNNGY 65
DB 42 YRWNAVIGGFTVGLVFPSPVGLAYARDIGGAYWDDRGARWPLDHLGDDWNL 101
QY 66 NGVSIADPINTNKWAAVGMVYNSWDPNPDGAILRSSDQATWQITPLPKLGNMGR 125
DB 102 LGVEAMAVDPHPRDLRLAVGTVAQSWAGN-GAVLRSEDGATWTRDITVKLGNEGR 160
QY 126 GMGERLAVDPNNILYFGAPSGKGLWRSTDSGATWSQMTNFPDVGTYIANPTDTTGVQS 185
DB 161 GAGERLIVDRSDSTLWLT-RHDGLLKSTDRGATWAAATAPP-----AKANS 207
QY 186 DIQGVVWAFDKSSSSLGQASKTIFVGVDPNP-----FVWSRDGGATWQAVGAPTGF 240
DB 208 SQGVWF-----IYAAGRTVYAGGDDGTSCTANLYRTAD-GTTWAGVGRSGT 257
QY 241 ---IPHKGVDPVNHVLIYATNSGTGPGYDGSVDWPKFVSFTGWTTRISPV-----PSTDT 293
DB 258 SAKVPLRAAYDTHRELIVVYDGPAGPGQSGSVHKLRTATGTWTEVTPVKPGGTTSDG 317
QY 294 ANDYFGVSLTIDRQHNPTMVAQISWPPITIFRSTDDGATWTRIDWTSYNRSLRY 353
DB 318 SADTFAYGVAVDARREGLTVNTRNWDGDTVFRSTDDGRTWTSKX-----AA 368
QY 354 VLDISAEPLTFGVQPPVPSPKLGWDEAMADPFNSDRMLYGTGATLYATNLTWTD 413
DB 369 VFDVSETPFLDWGDD-----KPKFGWMIQALAVDPYDSQHVYGTGATLYGTDLKRW- 421
QY 414 SGGQTHAPMVKGLEAVNDLISPPSG-APLISALGDLGFTHADVTAVPST-IFTSPV 471
DB 422 -----APRIRGLEEASVRLQISPPVGEAHLISGLGDIQVWYHERLTASTSRGMAINPV 474
```



OC	Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC	Xanthomonadaceae; Xanthomonas.
OX	NCBI_TaxID=340;
XP	[1]
RN	SEQUENCE FROM N.A.
RP	STRAIN=ATCC 33913 / NCPPB 528;
RC	MDLINE=22022145; PubMed=1202417;
RX	da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA	Quaggio L.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA	Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA	Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,
RA	Cicarelli R.M., Coutinho L.B., Cursino-Santos J.R., El-Dorri H.,
RA	Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA	Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA	Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA	Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA	Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA	Moreira L.M., Novo M.T.M., Okura Y.K., Oliveira M.C., Oliveira V.R.,
RA	Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA	Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA	Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA	Setubal J.C., Kitajima J.P.;
RT	"Comparison of the genomes of two Xanthomonas pathogens with differing
RT	host specificities.";
RL	Nature 417:459-463(2002).
DR	EMBL; AE012276; AAM41043.1; --
DR	GO; GO:0004197; F:cytosine-type endopeptidase activity; IEA.
DR	GO; GO:0006457; P:protein folding; IEA.
DR	GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR	InterPro; IPR001179; FKBP_PPase.
DR	InterPro; IPR002860; GH_BNR.
DR	Pfam; PF02012; BNR; 7.
DR	PROSITE; PS00453; FKBP_PPase_1; 1.
DR	PROSITE; PS00639; THIOU_PROTEASE_HIS; 1.
KW	Complete proteome.
SQ	SEQUENCE 751 AA; 81140 MW; 7378B6AA0B5698A4 CRC64;
Query Match	28.4%; Score 1145; DB 16; Length 751;
Best Local Similarity	35.08; Pred No. 2e-55;
Matches 266; Conservative 121; Mismatches 295; Indels 78; Gaps 21	
QY 1	ATTQPTWNSVAIGGGFVDGIVFENEGAPGLYVRTDIGMYRWDAAANGRIPLLDWVGW 60
DB 41	ATSGPYQWRSVAIGGGFVTGVLPHPAERGLAYARTDVGGAIRWDAAQQAQTALTDWLA 100
QY 61	NNVGYNGVSVIADPINTNKWAAVGYMTNSWDNDGAILRSSQGATWQITPLPFKLGG 120
DB 101	DDNLNMGIDAFADVPADADALYLAAGTYMHERAGN-AAVLSFRNRRGTFERADLPFKLGG 159
QY 121	NMPGRGMRGLAVDPNNDNITYLFGAPSGKGLWRSTDGATWSQMTNPDVGTIANPDTT 180
DB 160	NQLGRANGERLAVDPHGRVLLGLSRDA-GLWRSDDRGAHWKVASPPDAAALAGATARNH 218
QY 181	TGYQSDITQGVVWVAFDRSSSLGQASKTIFVGVADPNPNFWSRDGGATWQAVCAPGTF 240
DB 219	VGREQAV-GIAFVVFDAAASGNTGTPTRIYGVSTEQTSLSYVSEDAGRSWAPVAGPRL 277
QY 241	IPHKGVDPNVNVLVIATNSVTGGPYDGGSGGVWKFVSVTGWTTRISPVPTSTDTANDYFGY 300
DB 278	RPSHMAGSGDGH-WYLSYGGPQPGDLMAGGALWKFPTPAQGRWRBISPTPQ-PASGDGFGW 335
QY 301	SLGTIDRQHNITMVAIQISWMPDTIIFRSTDGATWTRIDWTSYPNRSRLRYVLIDISAE 360
DB 336	GAVAVDPQQQOVLLASTFRRTPRDELYRSVDGGKHAPL-----LADAVFDHSA 386
QY 361	PWLTFTGVPQNPVPSPKLGWDEMAAIDPFNSDRMLYGTGATLYATNDLTKWDSGGQTHI 420
DB 387	PW-TAHATPH-----WMG-ALAIADPFGNHAFVTGVIWASRL-----QDFA 428
QY 421	AP-----MWKLETAVNDLISPPSGAPLISALGDGLGFTHADVTAVPSTIFTSPVF 472
DB 429	APORPLWQDRLGEETVPLDLLSPMAGAHLLSALGIDIDCFRDELDRA-OLOVAGPRL 487

Qy	473	TTGTSVDYAEINFSIIIVRAGSFDPSSSQPNDRHVAFTSDGGKWNFP-QSEPGGVITGGTGA	533
Dd	488	TNGESIDAAQAQQVVRSGTV-RDRENNEIRALYSRDGGKWTAFASEPAPQGAGSIA	546
Qy	532	ASADGRFWWAGDGPPGPVVAVVGFGNSWAASCQPANAIQRSDRVNPKTYA--LSNGT	589
Dd	547	IGADAQQVWAPERGGN--WRTSDFGAQWQRVDGLPDTAVVMADRVDARRRYAYDVASGQ	604
Qy	590	FYESTDGGVTTFQPVAAAGL-----PSSGAVGVMFHAVPGKEGDLWLASS	633
Dd	605	LYESTDAARFR--ATGVQGSFARDERTRPQLRDPDWRAAGVTVLASPGK-----	652
Qy	634	GLYSTHTGGSSSAITGVSSAVNNVGFKSAPGSSYPAFVFVGTTCGGVTGAYRSDDCGTTW	693
Dd	653	GVMRWQDG--TLQVLSPDEARSILGICKALRAGAPPALYLAGRQGVGDVFRSDDGGVQM	710
Qy	694	VLNDOHQVGNWGQAITGBHANLRVYIQTNGRGIVYGD	733
Dd	711	QRINDDAHPRGR-PYSVTGDPRIAGRIFYATGCRGIFYG	749
RESULT 13			
ID	Q8J0D2	PRELIMINARY;	PRT; 812 AA.
AC	Q8J0D2;		
DT	01-MAR-2003 (TEMBUREl. 23, Created)		
DT	01-MAR-2003 (TEMBUREl. 23, Last sequence update)		
DT	01-JUN-2003 (TEMBUREl. 24, Last annotation update)		
DE	Oligoxylloglucan reducing end-specific cellobiohydrolase.		
OS	Geotrichum sp. M128.		
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
OC	Saccharomycetales; Dipodascaceae; mitosporic Dipodascaceae;		
OC	Geotrichum.		
OX	NCBI_Taxid=203496;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	MEDLINE=22359052; PubMed=12374797;		
RX	STRAIN=E22359052; PubMed=12374797;		
RA	Yaoi K., Mitsuiishi Y.;		
RT	"Purification, Characterization, Cloning, and Expression of a Novel		
RT	Xyloglucan-specific Glycosidase, Oligoxylloglucan Reducing End-specific		
RL	Cellobiohydrolase.";		
RL	J. Biol. Chem. 277:48276-48281(2002).		
DR	EMBL: AB089343; BAC22065.1 -		
DR	GO: GO:0016787; F:hydrolase activity; IEA.		
DR	InterPro: IPR002860; GH_BNR.		
DR	Pfam: PF02012; BNR; 9.		
KW	Hydrolase.		
SQ	SEQUENCE 812 AA; 87080 MW; P5C740F32DBB17E5 CRC64;		
Query Match 27.7%; Score 1117; DB 3; Length 812;			
Best Local Similarity 35.0%; Pred.No. 7.7e-54;			
Matches 281; Conservative 124; Mismatches 279; Indels 120; Gaps 31			
Qy	6	YTWSNVAICGGGFVDGIVFNVEGAPGILIYVRTDIGMYRWDAAANGWIPLLDVGMNNGY	65
Dd	27	YEFNVVAICGGGYTGIVAHPTKDLYARTDIGAYRWDACTSKNIPLNDFFAQDNVI	86
Qy	66	NGVVSIAADPIINTKNVAAVGMT-NSWDPNDGAILRSSDQCATWTQITPLPFKLGNMGP	124
Dd	87	MGTESIALDPNPDLRYLAQRYVGDEW---AAFVYVEDRGQSFTTIESPFPMGANDMG	142
Qy	125	RMGERLAVDPNDNLITYFGAPSGKLWRSTDSGATWSQMTNFPDVGTYYIANPTDTTGYQ	184
Dd	143	RNNGERLAVNPFFNSNEYMMGTTRT-EGIKWSDKRAKTNTVTSIPDAF-----TNIGIGYT	195
Qy	185	SDIQGVVWVAFDKSSSLGASKTIFFGVADPNPVPFWSRDGGATWAVPGAPTGI---	241
Dd	196	S-----VLTPD-----ERNGTIYASATAPOG-MYVTHDGGVSWEPVAGOPSSLNRT	241
Qy	242	-----PHKGVPDPVNHVLIYATNSTGGPYDGGSGDVWKFKSVTSQWTTRI-	285



```

Qy 484 NPSIIVRAGSFDPSQP-----NDRHVAFSTDGKKNWFQGS----- 519
Db 490 TPLV-----FSPTDPKTLVYGMQHVLRTNDGAVNTEISPDLTQKTTEPVKKNHADFDD 543
Qy 520 ---EPGGVTTG--GTVAASADGSRFVWAPGDPQPVVYVAVGFGNSW--AASQGVPAANAQI 572
Db 544 EEDKQPAPANGVIQTIAPSAKAGIIVWGTSTGLVQLTRDG-GKNWQNVTPPGLPERRAI 602
Qy 573 RSDRVNPK---TFYAL-----SNGTFYRSTDDGGVTQFQVAAAGLPSSGAVGVMPHAFVPGK 623
Db 603 ILIEASPRNADTAVIGVAPDSDHPYIYRTDGGKSWQKIVTGLPKGIARVV-REDPTR 661
Qy 624 EGDMLAASSGLYHSTNGGSSWSAI 648
Db 662 EGLVYAGTETGAHVSYDGDHWQTL 686

```

Search completed: May 11, 2004, 12:09:26  
Job time : 90.3468 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 12:04:16 ; Search time 36.3141 Seconds  
(without alignments)  
1052.023 Million cell updates/sec

Title: US-09-917-376-3

Perfect score: 4036

Sequence: 1 ATTQPTWSNVAIGGGFVD.....YIGTNGRGIYVIGGAPSG 740

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_AA.\*

1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*

2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*

3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*

4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*

5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*

6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	219.5	5.4	2736	4	US-09-252-991A-30227
2	181.5	4.5	3892	4	US-09-328-352-5503
3	162.5	4.0	772	2	US-08-410-784A-5
4	158.5	3.9	774	3	US-09-346-237-8
5	157.5	3.9	688	4	US-09-489-039A-9813
6	154	3.8	1548	4	US-09-252-991A-22301
7	152.5	3.8	1751	3	US-09-136-574A-44
8	152	3.8	3472	4	US-09-408-020-4
9	151.5	3.8	776	3	US-09-346-237-4
10	148.5	3.7	750	6	5457037-3
11	148.5	3.7	751	6	5457037-5
12	148.5	3.7	776	3	US-09-346-237-7
13	147	3.6	3623	4	US-09-341-461-2
14	146	3.6	1749	4	US-09-640-419C-28
15	143	3.5	894	3	US-08-362-525-22
16	143	3.5	894	3	US-08-971-692-15
17	142	3.5	1216	4	US-09-134-000C-5130
18	141	3.5	720	3	US-09-296-284-25
19	141	3.5	754	3	US-09-296-284-4
20	140.5	3.5	387	4	US-09-252-991A-21962
21	140.5	3.5	1000	3	US-09-352-159-25
22	140.5	3.5	1000	3	US-09-352-159-25
23	139	3.4	826	4	US-09-328-352-7515
24	137.5	3.4	1298	4	US-09-252-991A-30579
25	137	3.4	619	4	US-09-252-991A-26352
26	136	3.4	2628	2	US-08-570-311-14
27	135.5	3.4	1205	3	US-09-352-159-29

RESULT 1  
US-09-252-991A-30227  
; Sequence 30227, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 30227  
; LENGTH: 2736  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-30227

Query Match Similarity 5.4%; Score 219.5; DB 4; Length 2736;

Best Local Similarity 21.2%; Pred. No. 4.8e-07;  
Matches 190; Conservative 94; Mismatches 324; Indels 289; Gaps 48;

QY	59	GNWNNGY-----NGV--SIAADPI-NTNKVA-----AVGMYTNSWDPNDCGAILR-SS	103
Db	1305	CGSNWTFPTATPLANGTVVNAVAQDPAGNTGPGSTTVDAVAPNTFVNPSNGLNGTA	1364
QY	104	DQATWQITPLPKLGGNMPGRGMRGLAVDPNNNDILYFGAPSGKLMRSTDSGATWSQ	163
Db	1365	BERGSTVTLT-----DGN--GNPIGTTA-----DGSNGNSFTPG-----SQ	1398
QY	164	MTNFPDVGTIYI-ANTDTTGYQSDTCGVVWAFDKSSSLGQ---ASKTIFVGVDPNP	219
Db	1399	LPN-----GTVVNVTASDAAGNTS---APATTVTSSLPSPQVDPNSGVISGTADAGNT	1451
QY	220	VF-----WSDRGATWQAVPGAP-----TGFIIP	242
Db	1452	IIITDGNPNICQVTDAGSGNNSFTPGIPLDGTVVNVVARSNVDSPAPVITVDGVAP	1511
QY	243	HKGVPDPVNHVLIYIATSNITGGPY---DGSSGDVWFKEFSVT-SGTWTRISPVV	289
Db	1512	AAPVIDPSNGTISGTAEAGATVILTDGGNPIGQATADSGNWTFTPGTPLANGTVINA	1571
QY	290	-STDANDYFGVSGLTIDRQHNTIMVATQISWNPDTIIFRSTGCGATWTRIDWTSYPN	348
Db	1572	VAQDPAGNTSGPASTVTDIAIP-----PAPVINPSNGVIGSGTAEAGAT-VILTDGNGNP-	1625

## ALIGNMENTS





Db 340 DKTINTSFRGLONPTYSITADFQNSWNTVCGNTYNTNTIAQNLIIVSLAYWRDKL 399  
Qy 386 AIDPFNSDRMLYGCATLYATNDLTKWDSGGQIHIAPMVKLEETAVNDLIS--PP----- 439  
Db 400 GVDGYRDLASVIGNSCQHCFNFKMDAG-----NALNRIIVAEUPPRAT 445  
Qy 440 --SCAPLIS--ALG-----DILGF-----THADVTAVPSTIFTSPTTGTSDVYA 481  
Db 446 GSGVDLIAEPAIGNSYQVGGFPSCWAENWAGAYRDVVRQAOKLGSVAITTG----- 499  
Qy 482 ELNPSIIIVR-AGSDPPSSQPNDRHVAFTSDGKKNWFG--SEPGVTTGTGTVAAASDGS 537  
Db 500 ---QMATRFAGSSD-----LYGDDGRKPMHVSFNITAHGDFTLKDLYSCNSKKN 545  
Qy 538 RFVW--APGDPGQPVVYAVFGNSW-----AASQ-----GVP-----ANAQ 571  
Db 546 NQWPFYGPSDGE-----DNNNSWDQGGIAADQKARNGMALMLLSAGVPMIVGGDEA 599  
Qy 572 IRS-----DRVNPKTFY-----ALSNCTFYRSTD-G 596  
Db 600 LRSNMCNNPNYLDSSANLWNSRTTQNNFQSFASKAMIAFRKAHPALRPAFYSSVDNN 659  
Qy 597 GVTQPVAAAGLPSSGAVCMF-----HAVPGK-EGDLWLAASSGLYHSTNGGSSWSAIT 649  
Db 660 GNVNEQLRWFKPDGGVADATYFNDANNHAIAMRIDGSEFGDTASAIYVAHN---AWSA-- 714  
Qy 650 GVSSAVNVGFKSAPGSSYPAVFVVGTTGGVTVAYRSDCCTTWLINDDQHOYGNWQ 708  
Db 715 -----QVNFILPWFAGKSWYRVTDCGWAEGASQVQAPGE-ALVGENTAYGLCGR 766

RESULT 5  
US-09-489-039A-9813  
; Sequence 9813, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 9813  
; LENGTH: 688  
; TYPE: PRT  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-9813

Query Match 3.9%; Score 157.5; DB 4; Length 688;  
Best Local Similarity 22.6%; Pred. No. 0.0025;  
Matches 111; Conservative 53; Mismatches 184; Indels 143; Gaps 24;

Qy 60 WNNWG--YNGVVSTAAADPINTNKVAAVGMVYNSWDNDGAILRSSDQATWQITLPEFK 117  
Db 171 WAHGNTHGDRFAALDQINKQN-----VNQLQVAVAHGTGDIPOSNGSGAEDQNTPLQ-- 224  
Qy 118 LGGNMGRGMRGLAVDPNNDNIIYFGAPSGKGLWR--STDGATW-----SQM 164  
Db 225 -----IGDTLYVCTFPYSKVLADVDSCKRWYDSKSSPNWQRCRGLGYVADSQA 275  
Qy 165 TNFPDVGTT-----YIANPDT-----TGYSOI---QGVVWVAFD 196  
Db 276 QTAPASGTQPAACSRRLFLPTIDARLIAIDADTKLCENFGDGGIVDLISVGMGEVKAGY 335  
Qy 197 KSSSSLQASKTIIVG--VAD--PNNPVFWSRGGATWQAVPGAPTGFIPHKG-----VF 247  
Db 336 QQTSTPLVAGNVVVGVDADNYSTGEP-----PGVVRADFVHTGKLAWAW 381  
Qy 248 DPNVNHLYIATSNTPGPGYDSSGDWKFVSFTSGT---WTRIS-----PVPSTDTAND 296

## RESULT 6

US-09-252-991A-22301

; Sequence 22301, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 22301

; LENGTH: 1548

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-22301

Query Match 3.8%; Score 154; DB 4; Length 1548;

Best Local Similarity 20.7%; Pred. No. 0.014;

Matches 188; Conservative 87; Mismatches 270; Indels 364; Gaps 53;

Qy 73 ADPINTNKVAAVGMV---TNSWD-----PNDGAILR----- 101

Db 319 SDGVRQGRNWAELAEMLPAGSQSWSMRLVAGADLAAADNRLVRPDPSSASLADTHYQAKI 378

Qy 102 -----SSDQATWQITLPLFKLGGNMPGRGMRGLAVDPNNDNIIYFGAPSGKGLWR 153

Db 379 EQSSGLVFTQATDWGITP-----GTPVDESNEWI-----CGLGPLYC 416

Qy 154 STDGATWSQMTNFPDVGTYIANPTDTTGYQSDIQGVVWVAFDKS--SSSLQASKTIFV 211

Db 417 AEPPTWAP-----GNLCPAGTAGEGDLW---MCSVDPSLCIENLGT----- 460

Qy 212 GVADPNNPVF-----WSDRG--GATWQAVPGAPTGFIPHKGVFDP 249

Db 461 -VVTQNLQFLSVLTCTGDLDLASAGNLQWSPYGVYTAGTQAADVA--TCFNQPRGLFN- 517

Qy 250 VNHVL-----YIATS-----NTGGPYD---GSS--GDVWKFVSFTGWTTRISVPV 289

Db 518 -GSVLGAGADYEVLSTSQYQAWPEHGNLNDIAVGGDVVDQWAEKLTSS--DPIRPLP 574

Qy 290 STDANDYFGYSGLTIDRQHPNTIMVATQISWMPDTIIF-RSTDGATWTRIMDWT--- 344

Db 575 PSAAVGNLWLRQG--SADREGVFT-----AMWVFGSVRGAEGDAPY--LVGFTGFGT 624

Qy 345 -----SYPN-----RSRYVLDISAEPLT----- 364

Db 625 LGGGNLSMRGTGDAGNIAPRGDGSIPSSGNLNPSSQGLVLA VAGTGRLTSDGALQGGGG 684

Qy 365 -----FGVQNPVPSPKLGWDEAMDAIDPFNSDRM--LYGTG-----AT 402  
Db 685 DLNVRIGGEVNP-----SREARATQYSSGDFGLYSGGTIHDLOQALINLRGSAS 735  
Qy 403 LYATNDLTWKDSCGOIHAP-----MVKGLEETAVNDLISP-----PSGAPLIS-----446  
Db 736 LY-----SGALGGIDPRYDTLLRDPAEVRSRDAFSPTLASSTGGLTLVAGDTCMRL 786  
Qy 447 -ALGD--LGGFTRADVAVPSTI--FTSP-----VFTTGTSDVDAELNPSII 488  
Db 787 ETRGDVLVGGVDPGRVGVNTVGTFTAPDGSVYGGGIGHFSLWTAHTSID-----LP 839  
Qy 489 VRAGSFDPSQ-----PNDRHVAFTSD-----GGKNWFGSE 520  
Db 840 AAGNLFTQVLEATNAIPWAGNLSPSDGRFIYPSIVRAAAPEGSYILGPSSGYMGV 899  
Qy 521 PGVYTT---GGTVNASADGRFVWAPDGPQPVVYAVFGNSWAAASQGVPAQAQI-----572  
Db 900 SLNVSTTPYSLLAPSLNGELELLA-GDS-----IYAGGY--SVQSGADPANLPSIWTPA 952  
Qy 573 ---RSDR--VNPKTFYALSNGTFYRSTDDGGVTFQPVAAAGLPSSGAVGMFHAVPGKEGDL 627  
Db 953 FAGYSDAALNP-----IAGNG-----SPDGN---PAVIGGLP-----LFFGPDASAAL 994  
Qy 628 WLAASSGLYHSTNGSSWSAITGVSSAVNVGFGKSAPGSSYPVAVVGTIGGVGTAYRSD 687  
Db 995 ARDLQPARFYALTGD-----IVGLNSGAQIRFGEQA-----GNRAGQTWYE 1035  
Qy 688 DCGTTWVLINDHQYQN-WQAI-----TGD---HANLRVYIGTNGRGIYV 731  
Db 1036 GAGPVMWRAGRDIVASGTPLQRIAPSQISTDASFTGNLFPVHDDPNLDSLVAQGRDILY 1095  
Qy 732 GDIGGAPSG 740  
Db 1096 GNFNVAGPG 1104

RESULT 7

US-09-136-574A-44

Sequence 44, Application US/09136574A

Patent No. 6294366

GENERAL INFORMATION:

APPLICANT: Farrington, Graham K.

Anderson, Paige

Gibbs, Moreland

Bergquist, Peter

Daniels, Roy

Morgan, Hugh W.

Williams, Diane P.

TITLE OF INVENTION: Compositions and Methods for

Treating Cellulose Containing Fabrics Using Truncated

Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49

CORRESPONDENCE ADDRESS:

ADDRESSEE: Howson and Howson

STREET: Spring House Corporate Center, P.O. Box 457

CITY: Spring House

STATE: PA

COUNTRY: USA

ZIP: 19477

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM compatible

OPERATING SYSTEM: DOS

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/136.574A

FILING DATE: 19-Aug-1998

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/932,571

FILING DATE: September 19, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Bak, Mary E.

REGISTRATION NUMBER: 31,215

REFERENCE/DOCKET NUMBER: 1997US001/CIP

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-540-9200

TELEFAX: 215-540-5818

TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 44:

SEQUENCE CHARACTERISTICS:

LENGTH: 1751 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 44:

US-09-136-574A-44

Query Match 3.8%; Score 152.5; DB 3; Length 1751;

Best Local Similarity 18.7%; Pred. No. 0.021;

Matches 184; Conservative 101; Mismatches 349; Indels 351; Gaps 49;

Qy 6 YTWSNVAIGGGFVDGIVFNEGAPGILYVRTD-----IGMYRW-----D 45  
Db 233 YTAATGFTSGGFIIDLGW---AAWLYIATNDSSYLTKAEELMSEYANGTNTWTQCWDD 289  
Qy 46 AANGRIWIPLLDWGMNMGVNGVVSIAADPINTNKVAAVGM-YTNSWDPNDGAILRSSD 104  
Db 290 VRYGTLIMAKITKEL--YKGAVERNLDHWTDRITYTPKGMAYLTGM---GSLRYATT 343  
Qy 105 QG-----ATWQ-----ITPLPKLGNMPPGR-----GMGERLAVDPNNNDI 140  
Db 344 AAFACVYADVADSGCDNSNKKTKYLNPAKSIQIDYALGST--GRSFVVGFGTNPQPHHRN- 400  
Qy 141 LYFGAPSGKGLWRSTDSCATWSQMTNFPD-----VGTIANPTDTTGYOSDIQGVVW-- 192  
Db 401 -----AHSSWANSMKIPEYHRHILYALVGGPGSDSDNDIDITDYVONE 444  
Qy 193 VAFDKSSSLGQASKTIFGVADPNPNVFWSRDGGATQAVPGAPTGFPHKGVPDPVNH 252  
Db 445 VADYNAGIVGALAK-----MYQLYGEPID--DFKAIEFTND 481  
Qy 253 VLYIAT--SNTGGPYDSSGDVWKFVSFTSGTWTRISPVPSTDTAN--DYF-----GY 300  
Db 482 EIFVESKEFGNSQGP---NVTIVISYIYNRTGM---PPRVTDKLSPKYFIDLTIELIQAGY 534  
Qy 301 SGLTIDROHPNTIMVATQISWWPDIIIFRSTDGGATWTRIMDWTSYPNLSRVVL--DISA 359  
Db 535 S-----PDVVKVDT-----YYIEGKISGPYVWD---KNRNIYVVLVDFSG 572  
Qy 360 EPWLTFFGVQNPVPSPKLGW---MDEAMADPPNSDRMLYGTGATLYATNDLTWKWDSG 415  
Db 573 TK-----IYPGEVEHKKQAQFKISVPQGYPHDPTN--DPSYKGLTSQLEKNKYIAAYDNN 626  
Qy 416 QGIHIAPMVKGLEETAVNDLISP-----PSGAP-----443  
Db 627 N-----LVWGLEPGAATSTPAPTSTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 680  
Qy 444 -----LISALGDL-----GGFTHADVAV-----P 463  
Db 681 VLYKNNETSASTGSIRPWFKIVNGGSSVDLSRVKIRYWTYVDGKPSQSAVCDMAQIGAS 740  
Qy 464 STIFTSPTVFTTGT--VDYAEELNPSIIVRAGSPDPSSQPNDRHVAFTSDGKN-----514  
Db 741 NVTNFVKLSGSGVSGADY--YLEVGFSSGAGLOQPKDGTGDIQVRFNKNDSNYNQADDDWS 799  
Qy 515 WFO-----GSEPGVTTGGTVAASADGSRFVWAPGDP-----546  
Db 800 WLOSMNTYGENAKVLYVDGVLVWQEPGAPATPATPTATPTPTPTPTPTPTPTPTPTPTPTPT 859  
Qy 547 -----GQPVVYAVGFNGSNWAAASQGVPA-----NAQIRSDRVNPKTFYALSN-GT---589  
Db 860 VSATPTAPTASPVG--GSYWTPTSESYGALKWYANGNLSSTPNVLPK--IKIENVGTGA 916

```

590 Qy -----FYRSTDGVT-----FQP-----VAAGLPSSGA-----VG 614
917 Db VDLSRVKVRYWYTIIDGEATOSVASSINPAYIDVKFVLGANAGGADYYVEICFKSGAG 976
615 Qy VMFHAVPGEGDLWLAASSGLYHSTNGGSSWA-----ITGVSSAVNVGCKSAPGS 666
977 Db VLAAGOSTKEIRLSIKQSGSYNQSNIDYSVRSANSYIENEKVGYDDVLVWGREPGRNA 1036
667 Qy SYPAFVFGVTIGVGTGA-----YRSDDCGTT-----WVLINDDOHQYGNWGOAIT 711
1037 Db QIKVWVANGNLGSMTNVLPNKIENVTGTTAVDLSRVKVRYWYTIIDGEATQSVSVTSSIN 1096
712 Qy GDHANLRVYIGTNGRGI-VYGDIG 735
1097 Db PAYIDVKFVLGANAGGADYYVIRG 1121

```

## RESULT 8

```

US-09-408-020-4
; Sequence 4, Application US/09408020
; Patent No. 6632937
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCOBP.002A
; CURRENT APPLICATION NUMBER: US/09/408,020
; CURRENT FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: 60/102,294
; PRIOR FILING DATE: 1998-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 3472
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-09-408-020-4

```

Query Match	3.8%;	Score 152;	DB 4;	Length 3472;
Best Local Similarity	21.0%;	Pred. No. 0.058;		
Matches 172;	Conservative 100;	Mismatches 244;	Indels 304;	Gaps 50;
QY	96	DGAILRSSDQAGTQWQITP----	LPFKLGGNMPGRGMGERLAVDPNNDNILYFGAPSGKGL	151
DB	986	DGTVVFLSD-GPTGRVYPSLNPFDISSAAG-----	GFVIYVGVSDIDAFSA-DGRNM	1038
QY	152	WRSTDGATWSQMTNPP-DVG-----	TYIANP--TDTTGYQSDIOGVVWVAF	195
DB	1039	LVADETGGIHRYLARSPEYIGTDPIKSSLGEFVETPSAAPRVQDLAGIAFSHDGMIMLAA	1098	
QY	196	DKSSS-----	SIGQASKTTFVGVADNPNPVFWS-----	RDCGATQWQAVPG- 235
DB	1099	GGSGSVHRYSLPSPYAVSGAKYBETAMIG-GSPSGLEFSSDGLRMFVDPDAGSETAAVYGL	1157	
QY	236	-APTGFIPHKGVDPV-----	NHVLVIATSNTCGP-----	264
DB	1158	AAPYGI-----	GEAPLPPPLFLGVGAEATLSPDGRHLV-----	PGRPGLSQYSLPSTNL 1208
QY	265	-----	YDGSGD--VMKFVSVTGTWTRISVPVSPSTDAN-----	DYFGYSG----- 302
DB	1209	ELCAEPRGIDGGSCEDIYAIFE-SPKRGEGVSLAASITAAIDGPGICELGHGFAPPPMAPV	1567	
QY	303	---LTTDRQH-----	PNTIMVAT---QISWHPDT-----	IIIFR 329
DB	1268	MEQVTLDSREGTLRVLRDLRTVDVTRPYKMWVEDSDGSQTTLANSTLLNAENSNTLLIFR	1327	
QY	330	STDGGA-----	TWTRIW---DMTYSYNNR-----	SLRYVLIDISAEPLWTFGVQ 368
DB	1328	LDDAAAGKISGYTSPVFRTWSSFLGTGDGATRPHTLGFGDVLADIYDASGD-----	1379	
QY	369	PNPVPSPK-LGMWDEAMADPFNSDRMLYGTGA-----	TLVATNDLTKWDSGGQTHIA	421

```

Db 1380 -----VPSPSGIEFSDQMRM-----FTVGIGTGPGINIFTLSAPFDITLPHKSGSTNIG 1428
Qy 422 PMVKGLEETAVNDLISPPSGAPL-----ISALGD-----LGGFTH-----ADVT-A 461
Db 1429 -----GL-----SVSOLAFANNNGNSUTVLVDUGLVLEYALGDDYNVVTGTTQKFRITLDTTQG 1481
Qy 462 VPSTIETSP-----VFTTGTSDVYA---ELNPSIIIVRAGSFDPSS-----OP 500
Db 1482 IPNSIYTPDGLSQFVAYDDRLDLYVLGSPNDISSITTEIIPYSLPRP---DPTGMDFTF 1538
Qy 501 NDRHVAFSTDGGKMFQSGEPGGVTGG-----TVAASADGSRFV-----WAPGDPG-- 547
Db 1539 DGRMFLSTENGIDQYLLSEPPAVTTSVFLRTTIPIDGGAEGIRFVDNGRGLFVPGADGII 1598
Qy 548 --QPVVYAVGFGNSW--AASQGVPANQAIRSRDVRNPKTFYALSNCHGFYRSTDCGVTFQPV 603
Db 1599 QRHELIYPYGASTSLLETVRDGV-----TDGGPGENPA 1631
Qy 604 AAGLPSSGAVGVMFHAPVKGEDLWLAASGLYHSTNGSSWSAITGVSSAVNVGFKSA 663
Db 1632 AGSIRLAGT-----FNASDN-----VQSPSGIEFSGDGTGMF--VTGFGAAGVNEFSLSA 1679
Qy 664 P-GSSYPAPVFWGTIGG---VTGAYRSDDCGTTWVLINDD 699
Db 1680 PFDTTLVPHVELHDIGQPAVDLAFABD--GRTLLLLAAD 1717

RESULT 9
US-09-346-237-4
; Sequence 4, Application US/09346237A
; Patent No. 6265197
; GENERAL INFORMATION:
; APPLICANT: Bisgard-Frantzen, Henrik
; APPLICANT: Svendsen, Allan
; TITLE OF INVENTION: Starch Debranching Enzymes
; FILE REFERENCE: 5629.200-US
; CURRENT APPLICATION NUMBER: US/09/346,237A
; CURRENT FILING DATE: 1999-07-01
; EARLIER APPLICATION NUMBER: PA 1998 00868
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: 60/094,353
; EARLIER FILING DATE: 1998-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 776
; TYPE: PRT
; ORGANISM: Pseudomonas amyloideramosa
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(776)
; OTHER INFORMATION: Isoamylase
US-09-346-237-4

```

[illegible]

```

Qy 241 IPHKGVPDPVNHVLIATNTGGPYDGGSDGVWKFVSTGVTGTWTRISVPVST----- 291
Db 302 QAMVQAFHNAKIKYMDV-----VYNHTAEGGTWTSDDPTTATIIYSWRGLDN 348
Qy 292 -----DTANDYF-GYSLGTLIDRQHPNTI---MVATQISWPDFTIIFRSTGGGATWTRIW 341
Db 349 ATYYELTSGNQYFYDNTGIGANFNTYNTVAQNLI VDSLAYWANTM---GVDGFR-----F 400
Qy 342 DWTS-YPNRSLRYVLDISAEPLWLTGVPQNPVPSPKLGWMDAEMAD-----PENSDRML 396
Db 401 DLASVLGNSCLNGAYTASA-----PNCNGGYNFDDAADSNAVINRLREFTVRPAA 451
Qy 397 YGTGATLYATNDLTKWDSGGQIH-IAPMVKGLEE-----TAVNDLISPPSGAPLI 445
Db 452 GSGGLDLFA-----EPWALGNSYQLGGFPQGWSEWNLFRDSLRAQNEL-----GSMTI 502
Qy 446 SALGDLGGLGTHADVTAVPSTIFTSPVTTGTSVDYAEINPSIIVR-----AGS 493
Db 503 YVIQDANDFSGS-----SNLFQSSGRSPWNSINFIDVHDGMTLKDQVYSCNGANNQAWP 556
Qy 494 FDPSSQPNDRHVAFTDGGK---NWFQSPGPGVTTGGTV---AASADGSRF-VWAPGD 545
Db 557 YGPS-----DGGTSTNYSWDQMSAG---TGAADVQRRRAARTGMAFEMLSAGT 601
Qy 546 P-----GQPVVYAVGFGN-----SWAASQGVPAANAQIRSDR-----V 577
Db 602 PLMQGGDEYLATLQCNNAYNLDSSANWLTYSWTTDQ---SNFYTFAQRLIAFRKAHPAL 658
Qy 578 NPKTFYALNSGTFRSTDDGVTGVPVAAGLPSSGAVG-----VMFHAVPGKEGD 626
Db 659 RPSSWYSGSLTWY-----QP-----SGAVADSNYWNNTSNYAIAYAINGPS-- 700
Qy 627 LWLAASSGLYHSTNG-----GSSWSAIT-----GVSSAVNVGFKSAPG 665
Db 701 --LGDSNSIYVAYNGWSSSVTFLPAPPSTGTQWYRVTDTCWNNDGASTFV-----APG 751
Qy 666 SSYPAVFVVGIGVGTGAYRSDCGTTWVLINDDQHOYGNWQAI 710
Db 752 SE-----TLIGG-----AGTT-----YGCQGS 770

RESULT 10
5457037-3
; Patent No. 5457037
; APPLICANT: TOGNONI, ANGELO; CARRERA, PAOLO; CAMERINI, BARBARA;
; GALLI, GIULIANO; LUCCHESI, GIUSEPPE; GRANDI, GUIDO; DI GENNARO, CARLO
; TITLE OF INVENTION: CLONING OF THE GENE CODING THE ISOAMYLASE
; ENZYME AND ITS USE IN THE PRODUCTION OF SAID ENZYME
; NUMBER OF SEQUENCES: 7
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/1,797
; FILING DATE: 08-JAN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 749,621
; FILING DATE: 19-AUG-1991
; APPLICATION NUMBER: 224,114
; FILING DATE: 25-JUL-1988
; SEQ ID NO:3
; LENGTH: 750
5457037-3

Query Match 3.7%; Score 148.5; DB 6; Length 750;
Best Local Similarity 20.7%; Pred. No. 0.013;
Matches 171; Conservative 88; Mismatches 275; Indels 291; Gaps 47;

Qy 27 GAPGILYVRTDIGMYRWDANGRWIPLLDVWNNWNGWGSIAADPINTKNVWAAVG 86
Db 70 GITGAVY-----YGRWGPWNPVSNWKGSGAGFVSDVDANGDRFNPKNLLDYP 121
Qy 87 MYTNSWDPNDGAILRSSDQATWQITPLPFKLGGMFGRGMGERLAVDPNNNDIYFQAP 146
Db 122 AQEVSQDP-----LNPSNQ-----NGNVFASGASYR-----TTDSGIY--AP 156

```

```

Qy 147 SGKGLWRSTDSGATWSQMTNFPDV---GTVIANPTD-TTGYQSDIQGVVWVAFDKSS--SS 201
Db 157 KGVVLVPSTQSTGTFTPAQKDDVIYEVHVRGFTQDTSIPAQYRGTYGAGLKASYLAS 216
Qy 202 LGOASKTIFVGVAD-----PN---NPVFM-----SRDGGATWQAVPGATG-F 240
Db 217 LG-VTAVBFLPVQETQNDANDVVPNSDANQNTYGYWNTENYFSPDRRYAYNKAAGGPTAEF 275
Qy 241 IPHKGVPDPVNHVLIATNTGGPYDGGSDGVWKFVSTGVTGTWTRISVPVST----- 291
Db 276 QAMVQAFHNAKIKYMDV-----VYNHTAEGGTWTSDDPTTATIIYSWRGLDN 322
Qy 292 -----DTANDYF-GYSLGTLIDRQHPNTI---MVATQISWPDFTIIFRSTGGGATWTRIW 341
Db 323 TTYELTSGNQYFYDNTGIGANFNTYNTVAQNLI VDSLAYWANTM---GVDGFR-----F 374
Qy 342 DWTS-YPNRSLRYVLDISAEPLWLTGVPQNPVPSPKLGWMDAEMAD-----PENSDRML 396
Db 375 DLASVLGNSCLNGAYTASA-----PNCNGGYNFDDAADSNAVINRLREFTVRPAA 425
Qy 397 YGTGATLYATNDLTKWDSGGQIH-IAPMVKGLEE-----TAVNDLISPPSGAPLI 445
Db 426 GSGGLDLFA-----EPWALGNSYQLGGFPQGWSEWNLFRDSLRAQNEL-----GSMTI 476
Qy 446 SALGDLGGLGTHADVTAVPSTIFTSPVTTGTSVDYAEINPSIIVR-----AGS 493
Db 477 YVIQDANDFSGS-----SNLFQSSGRSPWNSINFIDVHDGMTLKDQVYSCNGANNQAWP 530
Qy 494 FDPSSQPNDRHVAFTDGGK---NWFQSPGPGVTTGGTV---AASADGSRF-VWAPGD 545
Db 531 YGPS-----DGGTSTNYSWDQMSAG---TGAADVQRRRAARTGMAFEMLSAGT 575
Qy 546 P-----GQPVVYAVGFGN-----SWAASQGVPAANAQIRSDR-----V 577
Db 576 PLMQGGDEYLATLQCNNAYNLDSSANWLTYSWTTDQ---SNFYTFAQRLIAFRKAHPAL 632
Qy 578 NPKTFYALNSGTFRSTDDGVTGVPVAAGLPSSGAVG-----VMFHAVPGKEGD 626
Db 633 RPSSWYSGSLTWY-----QP-----SGAVADSNYWNNTSNYAIAYAINGPS-- 674
Qy 627 LWLAASSGLYHSTNG-----GSSWSAIT-----GVSSAVNVGFKSAPG 665
Db 675 --LGDSNSIYVAYNGWSSSVTFLPAPPSTGTQWYRVTDTCWNNDGASTFV-----APG 725
Qy 666 SSYPAVFVVGIGVGTGAYRSDCGTTWVLINDDQHOYGNWQAI 710
Db 726 SE-----TLIGG-----AGTT-----YGCQGS 744

RESULT 11
5457037-5
; Patent No. 5457037
; APPLICANT: TOGNONI, ANGELO; CARRERA, PAOLO; CAMERINI, BARBARA;
; GALLI, GIULIANO; LUCCHESI, GIUSEPPE; GRANDI, GUIDO; DI GENNARO, CARLO
; TITLE OF INVENTION: CLONING OF THE GENE CODING THE ISOAMYLASE
; ENZYME AND ITS USE IN THE PRODUCTION OF SAID ENZYME
; NUMBER OF SEQUENCES: 7
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/1,797
; FILING DATE: 08-JAN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 749,621
; FILING DATE: 19-AUG-1991
; APPLICATION NUMBER: 224,114
; FILING DATE: 25-JUL-1988
; SEQ ID NO:5
; LENGTH: 751
5457037-5

Query Match 3.7%; Score 148.5; DB 6; Length 751;
Best Local Similarity 20.7%; Pred. No. 0.013;
Matches 171; Conservative 88; Mismatches 275; Indels 291; Gaps 47;

```

```

QY 27 GAPGILYVRTDIGMYRWDAAANGRWIPLLDWGMNMGVNGVVSIAADPINTNKVAAVG 86
Db 71 GITGAVY-----YGYRAGPNWPNYASNGKGSQAGFVSDVDANGDRFNENKLLDDPY 122
QY 87 MYTNSWDNDGAILRSSDOGATWQITPLPFKLGNNMGRGMBERLAVDPNNDNLIYFCAP 146
Db 123 AOEVSQDP-----LNPSNQ-----NGNVFASGASYR-----TTDSGIY--AP 157
QY 147 SKGGLWRSTDGATWSQMTNFPDV--GTIANPTD--TTGYQSDIQGVVWVAFDKSS--SS 201
Db 158 KGVLPVSTQSTGKTPTKRAQKDDVIYEVHVRGFTQDTSIPAQYRGTYYGAGLKASYLAS 217
QY 202 LGQASKTIIFGVAD-----PN--NPVFW-----SRDGGATWQAVPGAPTG-F 240
Db 218 LG-VTAVEFLPVQETQNDANDVVPNSDANQNYGMYTENYFSPDRRYAYNKAAGGPTAEF 276
QY 241 IPHKGVDPVNHVLYIATNTGPGYDSSGVWKFVSTGWTTRISVPST-----291
Db 277 QAMVQAFNAGIKVYMDV-----VYNHTAEGGTWSSDPTTATYISWRGLDN 323
QY 292 -----DTANDYF-GYSGLTIDRQHNTI---MVATOISWMPDTIIFRSTDGATWTRIW 341
Db 324 TTYIELTSGNQFYDNTGIGANFNTYNTVAQNLI VDSLAYWANTM---GVDGFR-----F 375
QY 342 DWTs-YPNRSRLRYVLDISAEPWLTFGVQPNPVPSPKLGWMDMAID----PFSNDRML 396
Db 376 DLASVLGNSCLNGAYTASA-----PNCNGGYNFDAADSNVAINRILREFTVRPAA 426
QY 397 YGTGATLYATNDLTWDSGGQIH-IAPMVKGLSE-----TAVNDLISPPSGAPLI 445
Db 427 GSGGLDLFA-----EPWAI GNSYQLGGFPQGWSEWNGLFRDSLROAQNEL-----GSMTI 477
QY 446 SALGDLGGFTHADVTAVPSTIFTSPVFTTCTSDVYAEINLPSIIVR-----AGS 493
Db 478 YVTQANDFSGS-----SNLFQSSGRSPWNSINFIDVHDGMLTKDVYSCNANNQAWP 531
QY 494 FDPSSQPNDRHVAFTDGGK-----NWFQSGPPGVTTGGTV---AASADGSRF-VWAPGD 545
Db 532 YGPS-----DGTSTNYSWDQMSAG---TGAADVORRAARTGMAFEMLSAGT 576
QY 546 P-----QGPVVYAVGFCN-----SWAASQGVPAQAIRSDR-----V 577
Db 577 PLMQGGDEYLRTLQCNNAYNLDSSANWLTYSWTTDQ---SNFYTPAQRLIAFRKAHPAL 633
QY 578 NPKTFYALNSGTFYRSTDGVTTFQPVAAAGLPSSGAVG-----VMFHAVPGKEGD 626
Db 634 RPSSWYSGSLTWY-----QP-----SGAVADSNYWNNTSNYAIAYINGPS-- 675
QY 627 LWLAASSGLYHSTNG-----GSSWSAIT-----GVSSAVNVVFGKSPAG 665
Db 676 --LGDSNSIYVAYNGWSSSVTFTLPAPPSGTQWYRVTDTCDWNDGASTFV-----APG 726
QY 666 SSYPAVFVGTIGGVTGAYRSDCGTTWVLINDDQHOYGNWGOAI 710
Db 727 SE-----TLIGG-----AGTT-----YGQCGQSL 745

```

## RESULT 12

```

US-09-346-237-7
; Sequence 7, Application US/09346237A
; Patent No. 6265197
; GENERAL INFORMATION:
; APPLICANT: Bisgard-Frantzen, Henrik
; APPLICANT: Svendsen, Allan
; TITLE OF INVENTION: Starch Debranching Enzymes
; FILE REFERENCE: 5629.200-US
; CURRENT APPLICATION NUMBER: US/09/346.237A
; CURRENT FILING DATE: 1999-07-01
; EARLIER APPLICATION NUMBER: PA 1998 00868
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: 60/094.353
; EARLIER FILING DATE: 1998-07-28

```

```

; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 776
; TYPE: PRT
; ORGANISM: Pseudomonas species SMP1
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(776)
; OTHER INFORMATION: Isoamylase
US-09-346-237-7

```

Query Match 3.7%; Score 148.5; DB 3; Length 776;

Best Local Similarity 20.7%; Pred. No. 0.013;

Matches 171; Conservative 88; Mismatches 275; Indels 291; Gaps 47;

```

QY 27 GAPGILYVRTDIGMYRWDAAANGRWIPLLDWGMNMGVNGVVSIAADPINTNKVAAVG 86
Db 96 GITGAVY-----YGYRAGPNWPNYASNGKGSQAGFVSDVDANGDRFNENKLLDDPY 147
QY 87 MYTNSWDNDGAILRSSDOGATWQITPLPFKLGNNMGRGMBERLAVDPNNDNLIYFCAP 146
Db 148 AOEVSQDP-----LNPSNQ-----NGNVFASGASYR-----TTDSGIY--AP 182
QY 147 SKGGLWRSTDGATWSQMTNFPDV--GTIANPTD--TTGYQSDIQGVVWVAFDKSS--SS 201
Db 183 KGVLPVSTQSTGKTPTKRAQKDDVIYEVHVRGFTQDTSIPAQYRGTYYGAGLKASYLAS 242
QY 202 LGQASKTIIFGVAD-----PN--NPVFW-----SRDGGATWQAVPGAPTG-F 240
Db 243 LG-VTAVEFLPVQETQNDANDVVPNSDANQNYGMYTENYFSPDRRYAYNKAAGGPTAEF 301
QY 241 IPHKGVDPVNHVLYIATNTGPGYDSSGVWKFVSTGWTTRISVPST-----291
Db 302 QAMVQAFNAGIKVYMDV-----VYNHTAEGGTWSSDPTTATYISWRGLDN 348
QY 292 -----DTANDYF-GYSGLTIDRQHNTI---MVATOISWMPDTIIFRSTDGATWTRIW 341
Db 349 TTYIELTSGNQFYDNTGIGANFNTYNTVAQNLI VDSLAYWANTM---GVDGFR-----F 400
QY 342 DWTs-YPNRSRLRYVLDISAEPWLTFGVQPNPVPSPKLGWMDMAID----PFSNDRML 396
Db 401 DLASVLGNSCLNGAYTASA-----PNCNGGYNFDAADSNVAINRILREFTVRPAA 451
QY 397 YGTGATLYATNDLTWDSGGQIH-IAPMVKGLSE-----TAVNDLISPPSGAPLI 445
Db 452 GSGGLDLFA-----EPWAI GNSYQLGGFPQGWSEWNGLFRDSLROAQNEL-----GSMTI 502
QY 446 SALGDLGGFTHADVTAVPSTIFTSPVFTTCTSDVYAEINLPSIIVR-----AGS 493
Db 503 YVTQANDFSGS-----SNLFQSSGRSPWNSINFIDVHDGMLTKDVYSCNANNQAWP 556
QY 494 FDPSSQPNDRHVAFTDGGK-----NWFQSGPPGVTTGGTV---AASADGSRF-VWAPGD 545
Db 557 YGPS-----DGTSTNYSWDQMSAG---TGAADVORRAARTGMAFEMLSAGT 601
QY 546 P-----QGPVVYAVGFCN-----SWAASQGVPAQAIRSDR-----V 577
Db 602 PLMQGGDEYLRTLQCNNAYNLDSSANWLTYSWTTDQ---SNFYTPAQRLIAFRKAHPAL 658
QY 578 NPKTFYALNSGTFYRSTDGVTTFQPVAAAGLPSSGAVG-----VMFHAVPGKEGD 626
Db 659 RPSSWYSGSLTWY-----QP-----SGAVADSNYWNNTSNYAIAYINGPS-- 700
QY 627 LWLAASSGLYHSTNG-----GSSWSAIT-----GVSSAVNVVFGKSPAG 665
Db 701 --LGDSNSIYVAYNGWSSSVTFTLPAPPSGTQWYRVTDTCDWNDGASTFV-----APG 751
QY 666 SSYPAVFVGTIGGVTGAYRSDCGTTWVLINDDQHOYGNWGOAI 710
Db 752 SE-----TLIGG-----AGTT-----YGQCGQSL 770

```



RESULT 13  
US-09-341-461-2  
; Sequence 2, Application US/09341461  
; Patent No. 6586389  
; GENERAL INFORMATION:  
; APPLICANT: Hammond, Timothy G.  
; TITLE OF INVENTION: Cubilin Protein, DNA Sequences Encoding Cubilin  
; TITLE OF INVENTION: and Uses Thereof  
; FILE REFERENCE: D6148  
; CURRENT APPLICATION NUMBER: US/09/341.461  
; PRIOR FILING DATE: 2000-07-20  
; PRIOR APPLICATION NUMBER: PCT/US99/01259  
; PRIOR FILING DATE: 1999-01-21  
; NUMBER OF SEQ ID NOS: 40  
; SEQ ID NO 2  
; LENGTH: 3623  
; TYPE: PRT  
; ORGANISM: rat  
; FEATURE:  
; OTHER INFORMATION: amino acid sequence of rat cubilin protein  
US-09-341-461-2  
Query Match 3.6%; Score 147; DB 4; Length 3623;  
Best Local Similarity 20.0%; Pred. No. 0.14;  
Matches 181; Conservative 79; Mismatches 301; Indels 344; Gaps 45;  
QY 6 YTWNSVAIGGGF-VGGIVNEGAGLTYRTDGGMYRWDANGRWIPLLDWGNNWG 64  
Db 1731 YSASTSCGGSFYLDGIFNSDYP-----ADYG-----PNAECV-WN--- 1767  
QY 65 YNGVWSTAADP-----INTNK-----VMAAVGMYTNSWDNDGA 98  
Db 1768 -----IASFGNRLQLSFLNLSLNCNKFVEIREGNATGGLIGRYCGSLPGN-- 1819  
QY 99 ILRSSQGAQTWQITPLPKLGNMPPGMRGLAVDPNNNLYFGAPSGKGLWRST--- 155  
Db 1820 -YSSAEGGSLW---VRFVSDGSGTGMGFQARFNKFNENNNIVGTGKIASPFMPGKYPY 1874  
QY 156 DSGATW-----SMTN-FPD-----VGTIANPTDIT 181  
Db 1875 NSNYKVVNVNDYGIIGRILEMDIEPTTNCFYDSLKIYDFDTGSLIGTYCG--TQTE 1932  
QY 182 GYSDIQGVVW-VAFDKSSSLGQASKTIFGVADPNPNVFWSDGGATWQAVPGAPTGF 240  
Db 1933 SFSSRNLYTFQSSDSVSGRGLLEWFAVDVSDSTPPI-----APGACGGF 1981  
QY 241 I-----PHKGVDPVNHVLYIATSNITGPGYDGGSDVWKFVSTGTWTRISFPVSTDTAND 296  
Db 1982 MVTGDTVPVGFSP-----GWPREVANG-----ADCIWIIYAP-DSTVELNI 2021  
QY 297 YFGYSLGIDTDRQHPNTIMVATQISWPDITIFRSTGGATWTRIDWTSYPNRSLRVLD 356  
Db 2022 -----LSLD-----IBPQSCNDKLVKDGDS-----D 2045  
QY 357 ISAEPLMTFCGVQNPVPSPKLGWMDMAIDPNNSDRMLYGTG----- 400  
Db 2046 LSPBLAVLCGSPGPIRS-----TGEYMI-RTSDTSVAGTGFNASFKSCGGYLAD 2099  
QY 401 -----ATLYATNDLTKWDSGGQIHIAPMVKGLEE-----TAVNDLI---S 437  
Db 2100 RGVITSKYPTDYLPLNLCNSGVLVQGLIATV-GFEQPFQIQNRDSFCQGDYLVLRNG 2158  
QY 438 PPSGAPLISALGDLGGTHADVTAVPTIIFTSPTVTTGTSVDVYAEALNPSIIVRAGSPDPS 497  
Db 2159 PDNGSPPLGSRNGRFGC---MYAPSTLFTS----- 2187  
QY 498 SQNDRHVAF---STDGKWFQSGEGVTTGTV-AAADSGSRFWAPGDCPQPVYA 553  
Db 2188 --GNEMFVQFISDSSNGGQFKIRYEAKSLACCGTVYIGDADSGYLTSPNYPAN---YP 2242  
QY 554 VFGNSWAASQGVPANAIQIR-----SDRWPKTFY-----ALSNGTFRSTDGGVTF 600

Db 2243 QGAEICWILEAPPGRSIOLOFEDQNIETPNCVSYLELRDGNANRLVSKLCOGT-- 2300  
QY 601 QPVAAGLPSSGAVGMFHAVFCKEGDMLAAASSGLY--HSTNGSGSSWSAITGVSSAVNVG 658  
Db 2301 -----LPGS-----WVSSRERYLKFGTGGSSYMGFKAKYSTASCG 2337  
QY 659 PGKAPGS-----SYPA-----VPVGTIGVTCAY-----RSDDCGTTWVLI 696  
Db 2338 GTVSGDSGVIESIGYPTLIFYANNVFCQMFIRLPGGYLTSLSPEDFNLOSPPGCTKDFVEI 2397  
QY 697 NDD-----QHXYGNMGQAITGDHANLRRVYIGNG-----RGIVVGDIG 735  
Db 2398 WENTSGRVLGRYCGNSTPSSVDTSSNVASVXFVTDGSGVTASGFRLOFKSSRQVCGDLG 2457  
QY 736 GAPSG 740  
Db 2458 G-PTG 2461  
RESULT 14  
US-09-640-419C-28  
; Sequence 28, Application US/09640419C  
; Patent No. 6630615  
; GENERAL INFORMATION:  
; APPLICANT: Bidney, Dennis L  
; APPLICANT: Crasta, Oswald R  
; APPLICANT: Hu, Xu  
; APPLICANT: Lu, Guihua  
; TITLE OF INVENTION: DEFENSE-RELATED SIGNALING GENES AND METHODS OF USE  
; FILE REFERENCE: 35718/199009 (5718-92)  
; CURRENT APPLICATION NUMBER: US/09/640.419C  
; CURRENT FILING DATE: 2000-08-17  
; PRIOR APPLICATION NUMBER: 60/149,656  
; PRIOR FILING DATE: 1999-08-18  
; PRIOR APPLICATION NUMBER: 60/206,405  
; PRIOR FILING DATE: 2000-05-23  
; NUMBER OF SEQ ID NOS: 28  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 28  
; LENGTH: 1749  
; TYPE: PRT  
; ORGANISM: Synechocystis PCC6803  
US-09-640-419C-28  
Query Match 3.6%; Score 146; DB 4; Length 1749;  
Best Local Similarity 21.8%; Pred. No. 0.062;  
Matches 160; Conservative 81; Mismatches 306; Indels 188; Gaps 35;  
QY 119 GGN-----MPGRGM--GERLAVDPNNNLYFGAPSGKGLWRSTDSGATWSQMTNPPDVG 171  
Db 119 GGNANLFLMNPAGMIFGNASINVPDGFVTTGSAIGFG-----NDQMFQVFSNDYN 171  
QY 172 TYIANPTDITGYSQSDIQGVVWVAFKSSSL-----GQASKT-----IFGVADPNPNVFW 222  
Db 172 ALIGNPSQ-----FAFLANPGLIINAGDLSVTEGKNLTFLAGNIVNTGSLA 218  
QY 223 SRDGGATWQAVPGAPTGFIPHKG-----VFPDYNH-----VLYIATSNITGPDYD--- 266  
Db 219 APGGNITVAAPGQNRIRISQAGSLLSLEVEVSPQMNQGGSPFVLDPLTLTQAGSNL 278  
QY 267 -----GSSGDVWKFVSTGTWTRISPVSTDTANDYFGYSLGIDTDRQHPNTIMVA-TQISW 321  
Db 279 GLAVQPNQSV-----TTNGTNALVSLPGSVTISGNVDASGKSTN1SSGQVVAIGQIAV 334  
QY 322 WPDITIFRSTDGGATWTRIDWTSYPNRSLRVLDISAEPLMTFCGVQNPVPSPKLGW 381  
Db 335 QGATVDSVNGGGGT-----VRIGGDFQGO--LT-----LPNASQTLI 370  
QY 382 DEAMAIQPNFSDRMLYGTGATLYA-TNDLTKWDS-----GGQIHIAPMVKGLEET- 430  
Db 371 DSNVW---KADALLTGNGGTIVVWADDSRFSNISAQGTMGNG-----GGFVETS 420  
QY 431 -AVNDLISPPSGAPLISALGDLGGF-----THADVTAVPSTIFTSPVTTGTSVD 479

Db 421 GAKSLMVDVDTARVNTFATMGLBGLTWLLDPLBIIIVGTDDLLADPKLV---SVLTITTSLD 477  
Qy 480 ----YAEINPGLIIRAG-SFDPSSOPN-----DRHVAFTDGGKMWFGSBEPPGVTT 526  
Db 478 NGNVILQADQSIQVQAFSADPSAGNLTFTSPITTDALSLGTGSIIFANTGP--INT 535  
Qy 527 GGTVAAS-----ADGSRFVWAPGDPGPVYVYVAGFGNSWAASQGVPAANAQI-- 572  
Db 536 GNTLVTSPTNLDPNKIQLNANTFTTAPG-----YDIYFRKSVNGGFDLLGNANFVY 588  
Qy 573 ---RSDRVNPKTFVALSNGTYRSTD----GGVTFQPVAAGLP-----SSGAVGMFHAVP 621  
Db 589 FDDGAGITPLKSFQVATETIYVGNNDIVTQGNQIFDGVFYGLQPVNLTSSAGSVIFTNNI 648  
Qy 622 GKEGDLWLAASSGLVHSTNGSSSAITGVSSAV-----NVGFGK-----SAPGS 666  
Db 649 LLNSGLQVQTAQIV--SQPSSLSAVEIADVLLNAGQVSGNINTRGGNVDIQLGN 706  
Qy 667 SYPAVFVVGTTGGVYGAYRSDCGTTWVLYLNDQHQYGNWQCAITGDHANLRVYIGTN- 725  
Db 707 ISTGSIVTSPFGNAGNVILNAGGT--LTTGVYETSGTNGGDVTTSSGSNTSTAYIDTRG 764  
Qy 726 -GRGIVYGDIGCAPS 739  
Db 765 FGDGLEIDSLGGAVS 779

RESULT 15

US-08-362-525-22  
; Sequence 22, Application US/08362525  
; Patent No. 6027910  
; GENERAL INFORMATION:  
; APPLICANT: KLIS, FRANCISCUS M.  
; APPLICANT: SCHREUDER, MAARTEN P.  
; APPLICANT: TOSCHKA, HOLSER Y.  
; APPLICANT: VERRIPS, CORNELIS T.  
; TITLE OF INVENTION: PROCESS FOR IMMOBILIZING ENZYMES TO THE  
; TITLE OF INVENTION: CELL WALL OF A MICROBIAL CELL BY PRODUCING A FUSION  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.  
; STREET: 1100 New York Avenue, N.W.  
; CITY: Washington  
; STATE: D. C.  
; COUNTRY: U.S.A.  
; ZIP: 20005-3918  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/362,525  
; FILING DATE: 04-JAN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EP 92202080.5  
; FILING DATE: 08-JUL-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EP 92203899.7  
; FILING DATE: 14-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/EP93/01763  
; FILING DATE: 07-JUL-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: KOKULIS, PAUL N.  
; REGISTRATION NUMBER: 16,773  
; REFERENCE/DOCKET NUMBER: 213289/T7020(V)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 861-3000  
; TELEFAX: (202) 822-0944

; TELEX: 6714627 CUSH  
; INFORMATION FOR SEQ ID NO: 22:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 894 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-362-525-22  
  
Query Match 3.5%; Score 143; DB 3; Length 894;  
Best Local Similarity 20.5%; Pred No. 0.041;  
Matches 161; Conservative 94; Mismatches 308; Indels 224; Gaps 39;  
  
Qy 7 TWSNVAIGGGFVDGIVFNEGAPGLIYVRTDTCGMWRDAANGRWIPLDWDVG----- 59  
Db 53 TVSNAAYMAYGA-----SKTKLGSVGGQTDISIDYN-----IPCSSSGTFFPCPQE 99  
Qy 60 ---WNRWNGVGVSIADPINTNKV---WAA--VGMVNSWD----- 93  
Db 100 DSYGNWCKGKGACG-----NSQGIAYWSTDLEGFYTTPTNVTLEMTGYELPQOTGSYTFK 155  
Qy 94 ---PNDGAILRSSDOGATW-----QITPLPFKLGGNMPGGRGERLAVDPNNDNIL 141  
Db 156 FATVDDSAI---SVGGATAFNCCAOQPITSTNFTIDGIKPWGG-----SLPNIEGTV 208  
Qy 142 YFGA----PSGKGLWRSTDGATWSQMTNFPDVGTVIANPTDTTGYQSDIQGVVWVAFDK 197  
Db 209 YWAGYVYPM-KVYVSNVSWGTLPISVTLPD-GITVSD-----DFEGVYV-SFD- 255  
Qy 198 SSSSLGQASKTIFVGVADPNPNVFWSRDGGATWQAVPGAPTGI PHKGVPDPVNVHLYIA 257  
Db 256 ---DDLQSQNC-----VPDSPN-----YAVSITTTTTTPEWTGFTSTSTEMTV 297  
Qy 258 TSNTGGPYD-----GSSGDVKNKFSVTSGTWTRISVPVSTDTANDYFGYSLTID-- 306  
Db 298 TGTNGVPTDETIVIRTPTSEGLI---STTPEWTGFTSTSTSEVTT-ITGNGQPTDET 353  
Qy 307 ---RQHPNTIMVATQISWPDITIFRSDGATWTRIDWTSYPNRSRLYVL----- 355  
Db 354 VIVIRTPTSEGLISITTEPTGTFTSTSE-----MTVTGTNGQPTDEIVIRTPTSEG 409  
Qy 356 ---DISAEPWL-TFGVQPNPFPVSPKLGWMDAEMADPFNSDRMLYGTGATLYATNDLTKW 412  
Db 410 LVTTTTEPTWGTG-----TSTSTEMSTVTGNGLPTDET-- 444  
Qy 413 DSGGQIHAPMYKLEETAUNDLISPPSCA-----PLISALGDLGGFTHADVAV 462  
Db 445 -----IVKTFPTTAISSLSLSSSSSGQITSSITSRPIITPFYPNG-TSVISSV 493  
Qy 463 PSTIETSPVFTTGTSDVVAELNPSIIVRAGSFDPSQPNDRHVAFTDGGKMWFGSBEPPG 522  
Db 494 ISSSVTSSLFTSSPVISSSVISSSTTTSISESKSSSVIPTSSTSSSE-SETSSAG 552  
Qy 523 GVTGTGTVAAADGSRFVWAPGDPGPVYVYVAGFGNSWAASQGVPAANAQIRSRVNPKT 582  
Db 553 SVSSSFISSESSKGP---TYSSSSLPVTSATTSQETASSL-PPATTTKTSQTTTLV- 607  
Qy 583 YALSNGTFYRSTDGVTTFQPVAAAGLPSSCAVGMFHAVPG--KEGDLWLAAS- 633  
Db 608 -----VTSCESHVCTESISPAIVSTATV-----TVSGVTTEYTTWCPISTTTKOT 654  
Qy 634 -GLYHSTNGGSSWSAITGVSSAVNVYFGKSPAGSSYPVAF--VVGITGGVTGAYRSDCG 690  
Db 655 KGTTEQTTETTKQTTVVTLISSCESDVCSKTAS-----PAIVSISTATINGVTTEY----- 704  
Qy 691 TTWVLIN 697  
Db 705 TTWCPIIS 711

Search completed: May 11, 2004, 12:11:20  
Job time : 40.3141 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 12:09:36 ; Search time 86.3468 Seconds  
(without alignments)  
2378.773 Million cell updates/sec

Title: US-09-917-376-3

Perfect score: 4036

Sequence: 1 ATTQPYTWSNVAIGGGFVD.....YIGTNGRGIVYDGGAPSG 740

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1140673 seqs, 277566755 residues

Total number of hits satisfying chosen parameters: 1140673

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4036	100.0	740	12	US-09-917-376-3
2	4036	100.0	740	12	US-09-917-376-6
3	4036	100.0	740	14	US-10-155-400-3
4	4036	100.0	740	14	US-10-155-400-6
5	4036	100.0	957	12	US-09-917-376-1
6	4036	100.0	957	14	US-10-155-400-1
7	2429.5	60.2	882	14	US-10-156-761-9395
8	1680	41.6	726	12	US-09-917-376-7
9	1680	41.6	726	14	US-10-155-400-7
10	1579	39.1	838	12	US-10-420-191-2
11	1573.5	39.0	818	14	US-10-026-994-2
12	1405	34.8	739	14	US-10-156-761-10111
13	1117	27.7	789	12	US-10-395-241-14
14	1117	27.7	812	12	US-10-395-241-12
15	1114	27.6	826	12	US-10-395-241-18

16	841.5	20.8	555	10	US-09-927-827-47	Sequence 47, Appl
17	238	5.9	2468	12	US-10-282-122A-66335	Sequence 66335, A
18	238	5.9	2468	14	US-10-246-330-4	Sequence 4, Appl
19	199.5	4.9	2435	12	US-10-282-122A-47453	Sequence 47453, A
20	193	4.8	1465	15	US-10-369-493-13955	Sequence 13955, A
21	181	4.5	1308	12	US-10-282-122A-44566	Sequence 44566, A
22	172.5	4.3	783	15	US-10-369-493-14105	Sequence 14105, A
23	170	4.2	1074	12	US-10-282-122A-50816	Sequence 50816, A
24	168.5	4.2	1439	12	US-10-282-122A-44567	Sequence 44567, A
25	167.5	4.2	1119	14	US-10-245-802-12	Sequence 12, Appl
26	165.5	4.1	613	15	US-10-369-493-11181	Sequence 11181, A
27	163.5	4.1	1289	12	US-10-282-122A-44901	Sequence 44901, A
28	158.5	3.9	774	9	US-09-833-435A-8	Sequence 8, Appl
29	158.5	3.9	774	14	US-10-375-720-8	Sequence 8, Appl
30	157	3.9	350	14	US-10-156-761-9225	Sequence 9225, Ap
31	156.5	3.9	2204	12	US-10-282-122A-64364	Sequence 64364, A
32	156	3.9	2117	15	US-10-120-801-63	Sequence 63, Appl
33	152	3.8	827	12	US-10-282-122A-47312	Sequence 47312, A
34	152	3.8	1721	12	US-10-282-122A-62548	Sequence 62548, A
35	152	3.8	3472	13	US-10-027-806-4	Sequence 4, Appl
36	152	3.8	3472	13	US-10-034-623-4	Sequence 4, Appl
37	152	3.8	3472	14	US-10-027-801-4	Sequence 4, Appl
38	152	3.8	3472	14	US-10-029-120-4	Sequence 4, Appl
39	151.5	3.8	776	9	US-09-833-435A-4	Sequence 4, Appl
40	151.5	3.8	776	14	US-10-375-720-4	Sequence 4, Appl
41	151	3.7	1291	14	US-10-156-761-14161	Sequence 14161, A
42	151	3.7	1325	12	US-10-282-122A-44953	Sequence 44953, A
43	151	3.7	1797	15	US-10-369-493-5176	Sequence 5176, Ap
44	151	3.7	1805	15	US-10-369-493-5177	Sequence 5177, Ap
45	150.5	3.7	1082	15	US-10-369-493-17890	Sequence 17890, A

#### ALIGNMENTS

#### RESULT 1

US-09-917-376-3  
; Sequence 3, Application US/09917376  
; Publication No. US2004003834A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOYTICUS  
; FILE REFERENCE: 40197.4US01  
; CURRENT APPLICATION NUMBER: US/09/917,376  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 740  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
; FEATURE:  
; OTHER INFORMATION: Catalytic domain GH74  
US-09-917-376-3

Query Match	100.0%;	Score 4036;	DB 12;	Length 740;
Best Local Similarity	100.0%;	Pred. No. 2.2e-301;		
Matches 740;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	ATTQPYTWSNVAIGGGFVDGIVFNEGAPGILYVTRDTCGMWRDANGRWIPLDVGW	60	
Db	1	ATTQPYTWSNVAIGGGFVDGIVFNEGAPGILYVTRDTCGMWRDANGRWIPLDVGW	60	
QY	61	NNWGVGVSTAADPINTNKVAAAGWYTNWDNDGALLRSSDQATQWITPLPFKLG	120	
Db	61	NNWGVGVSTAADPINTNKVAAAGWYTNWDNDGALLRSSDQATQWITPLPFKLG	120	
QY	121	NMPGRGMGERLAVDPNNDNIFYFGAPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDT	180	

Db 121 NMPGRGMRERLAVDPNNDNLLYFGAPSGKGLWRSTDSGATWSQMTNFPDVGTYIANPTDT 180  
Qy 181 TGYQSDIQGVVWVAFDPSKSSSLGQASKTIFVGVDAPNPNPFWSRDGGATWQAVPGATGF 240  
Db 181 TGYQSDIQGVVWVAFDPSKSSSLGQASKTIFVGVDAPNPNPFWSRDGGATWQAVPGATGF 240  
Qy 241 IPHKGVPDPVNHVLYIATNTGGPYDGSGLWKFVSCTGTWTRISPVPSDTDANDYFGY 300  
Db 241 IPHKGVPDPVNHVLYIATNTGGPYDGSGLWKFVSCTGTWTRISPVPSDTDANDYFGY 300  
Qy 301 SGLTIDRHPNTIMVATOISWMPDITIIFRSTDGGATWTRIDWTSYPNRSIRYVLDISAE 360  
Db 301 SGLTIDRHPNTIMVATOISWMPDITIIFRSTDGGATWTRIDWTSYPNRSIRYVLDISAE 360  
Qy 361 PWTFTGVQPNPVPSPKLGWDEAMAIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420  
Db 361 PWTFTGVQPNPVPSPKLGWDEAMAIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420  
Qy 421 APWKGLBETAVNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSPTTGTSDY 480  
Db 421 APWKGLBETAVNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSPTTGTSDY 480  
Qy 481 AELNPSIIVRAGSDPSSQPNDRHVAFSTDDGKNWFGSGEPGVTGTTGTTVAASADGSRFV 540  
Db 481 AELNPSIIVRAGSDPSSQPNDRHVAFSTDDGKNWFGSGEPGVTGTTGTTVAASADGSRFV 540  
Qy 541 WAPGDPQGVVYAVGFGNSWAASQGVPAQAQIRSDRVNPKTFYALNSGTFFRSTDGGVTF 600  
Db 541 WAPGDPQGVVYAVGFGNSWAASQGVPAQAQIRSDRVNPKTFYALNSGTFFRSTDGGVTF 600  
Qy 601 QPVAAGLPSSGAVGVMFHAPVKGEGDLWLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG 660  
Db 601 QPVAAGLPSSGAVGVMFHAPVKGEGDLWLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG 660  
Qy 661 KSPAGSSYPVAVVGTIGGVTGAYRSDDCGTTWVLINDDQHOYGNWQAITGDHANLRV 720  
Db 661 KSPAGSSYPVAVVGTIGGVTGAYRSDDCGTTWVLINDDQHOYGNWQAITGDHANLRV 720  
Qy 721 YIGTNGRGIVYGDIGGAPSG 740  
Db 721 YIGTNGRGIVYGDIGGAPSG 740

## RESULT 2

US-09-917-376-6

; Sequence 6, Application US/09917376

; Publication No. US20040038334A1

; GENERAL INFORMATION:

; APPLICANT: DING, SHI-YOU

; APPLICANT: ADNEY, WILLIAM S.

; APPLICANT: VINZANT, TODD B.

; APPLICANT: HIMMEL, MICHAEL E.

; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS

; TITLE OF INVENTION: CELLULOXYLICUS

; FILE REFERENCE: 40197.4US01

; CURRENT APPLICATION NUMBER: US/09/917,376

; CURRENT FILING DATE: 2001-07-28

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: Patent In Ver. 2.1

; SEQ ID NO 6

; LENGTH: 740

; TYPE: PRT

; ORGANISM: Acidothermus cellulolyticus

US-09-917-376-6

Query Match

Best Local Similarity 100.0%; Score 4036; DB 12; Length 740;

Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTQPTWSNVAICGGFVDGIVNEGAPGLIYRTDIDGMYRWDAANGRWIPLLDVWGV 60

Db 1 ATTQPTWSNVAICGGFVDGIVNEGAPGLIYRTDIDGMYRWDAANGRWIPLLDVWGV 60

Query Match 100.0%; Score 4036; DB 14; Length 740;  
Best Local Similarity 100.0%; Pred. No. 2.2e-301;  
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
; SEQ ID NO 6  
; LENGTH: 740  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulosoliticus  
US-10-155-400-6

QY 1 ATTPYTWNSNAIAGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 60  
DB 1 ATTPYTWNSNAIAGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 60  
QY 61 NNWGYGVVSIADPINTNKVAAVGMVYNSWDPNDGAILRSSDQATWQITPLPFLKGG 120  
DB 61 NNWGYGVVSIADPINTNKVAAVGMVYNSWDPNDGAILRSSDQATWQITPLPFLKGG 120  
QY 121 NMPGRGGERLAVDPNNNDILYFAGPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDT 180  
DB 121 NMPGRGGERLAVDPNNNDILYFAGPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDT 180  
QY 181 TGYQSDIQGVVWVAFDKSSSLGQASKTI FVGADPNPNPVFWSRDGATWQAVPGAPTGF 240  
DB 181 TGYQSDIQGVVWVAFDKSSSLGQASKTI FVGADPNPNPVFWSRDGATWQAVPGAPTGF 240  
QY 241 IPHKGVPDPVNHVLYIATNTGGPYDSSGVDWKFVSVTGWTWTRISPVPSDTDTANDYFGY 300  
DB 241 IPHKGVPDPVNHVLYIATNTGGPYDSSGVDWKFVSVTGWTWTRISPVPSDTDTANDYFGY 300  
QY 301 SGLTIDRQHPNTIMVATQISWMPDTIIFRSTDGATWTRIMDWTSPYNSRLRYVLDISAE 360  
DB 301 SGLTIDRQHPNTIMVATQISWMPDTIIFRSTDGATWTRIMDWTSPYNSRLRYVLDISAE 360  
QY 361 PWTFTGVPQNPVPSPKLGWDEMAIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420  
DB 361 PWTFTGVPQNPVPSPKLGWDEMAIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420  
QY 421 APYKGLTEETAANDLISPPSGAPLISALGDLGGFTHADVTAVPSTIIFTSVFTTGTSDY 480  
DB 421 APYKGLTEETAANDLISPPSGAPLISALGDLGGFTHADVTAVPSTIIFTSVFTTGTSDY 480  
QY 481 AELNPSIIVRAGSDPPSSQPNDRHVAFTDGGKWFQSGEPGVTGGTVAASADGSRFV 540  
DB 481 AELNPSIIVRAGSDPPSSQPNDRHVAFTDGGKWFQSGEPGVTGGTVAASADGSRFV 540  
QY 541 WAPGDPQPVVYAVFGNSWAASQGVPAQAIRSDRVNPKTFYALSNGTFYRSTDGVTFF 600  
DB 541 WAPGDPQPVVYAVFGNSWAASQGVPAQAIRSDRVNPKTFYALSNGTFYRSTDGVTFF 600  
QY 601 QPVAAGLPSSGAVGMFHAVPGKEDLWLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG 660  
DB 601 QPVAAGLPSSGAVGMFHAVPGKEDLWLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG 660  
QY 661 KSAFGSSYPAVFVGTIGVGTGAYRSDDCGTTWVLINDDQHOYGNWQQAITGDHANLRRV 720  
DB 661 KSAFGSSYPAVFVGTIGVGTGAYRSDDCGTTWVLINDDQHOYGNWQQAITGDHANLRRV 720  
QY 721 YIGTNGRGIVYDGGAPSG 740  
DB 721 YIGTNGRGIVYDGGAPSG 740

## RESULT 4

US-10-155-400-6  
; Sequence 6, Application US/10155400  
; Publication No. US20030108988A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS  
; FILE REFERENCE: NREL 01-36A  
; CURRENT APPLICATION NUMBER: US/10/155,400  
; CURRENT FILING DATE: 2002-10-22  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1

## RESULT 5

US-09-917-376-1  
; Sequence 1, Application US/09917376  
; Publication No. US20040038334A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS

Query Match 100.0%; Score 4036; DB 14; Length 740;  
Best Local Similarity 100.0%; Pred. No. 2.2e-301;  
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTPYTWNSNAIAGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 60  
DB 1 ATTPYTWNSNAIAGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 60  
QY 61 NNWGYGVVSIADPINTNKVAAVGMVYNSWDPNDGAILRSSDQATWQITPLPFLKGG 120  
DB 61 NNWGYGVVSIADPINTNKVAAVGMVYNSWDPNDGAILRSSDQATWQITPLPFLKGG 120  
QY 121 NMPGRGGERLAVDPNNNDILYFAGPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDT 180  
DB 121 NMPGRGGERLAVDPNNNDILYFAGPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDT 180  
QY 181 TGYQSDIQGVVWVAFDKSSSLGQASKTI FVGADPNPNPVFWSRDGATWQAVPGAPTGF 240  
DB 181 TGYQSDIQGVVWVAFDKSSSLGQASKTI FVGADPNPNPVFWSRDGATWQAVPGAPTGF 240  
QY 241 IPHKGVPDPVNHVLYIATNTGGPYDSSGVDWKFVSVTGWTWTRISPVPSDTDTANDYFGY 300  
DB 241 IPHKGVPDPVNHVLYIATNTGGPYDSSGVDWKFVSVTGWTWTRISPVPSDTDTANDYFGY 300  
QY 301 SGLTIDRQHPNTIMVATQISWMPDTIIFRSTDGATWTRIMDWTSPYNSRLRYVLDISAE 360  
DB 301 SGLTIDRQHPNTIMVATQISWMPDTIIFRSTDGATWTRIMDWTSPYNSRLRYVLDISAE 360  
QY 361 PWTFTGVPQNPVPSPKLGWDEMAIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420  
DB 361 PWTFTGVPQNPVPSPKLGWDEMAIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420  
QY 421 APYKGLTEETAANDLISPPSGAPLISALGDLGGFTHADVTAVPSTIIFTSVFTTGTSDY 480  
DB 421 APYKGLTEETAANDLISPPSGAPLISALGDLGGFTHADVTAVPSTIIFTSVFTTGTSDY 480  
QY 481 AELNPSIIVRAGSDPPSSQPNDRHVAFTDGGKWFQSGEPGVTGGTVAASADGSRFV 540  
DB 481 AELNPSIIVRAGSDPPSSQPNDRHVAFTDGGKWFQSGEPGVTGGTVAASADGSRFV 540  
QY 541 WAPGDPQPVVYAVFGNSWAASQGVPAQAIRSDRVNPKTFYALSNGTFYRSTDGVTFF 600  
DB 541 WAPGDPQPVVYAVFGNSWAASQGVPAQAIRSDRVNPKTFYALSNGTFYRSTDGVTFF 600  
QY 601 QPVAAGLPSSGAVGMFHAVPGKEDLWLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG 660  
DB 601 QPVAAGLPSSGAVGMFHAVPGKEDLWLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG 660  
QY 661 KSAFGSSYPAVFVGTIGVGTGAYRSDDCGTTWVLINDDQHOYGNWQQAITGDHANLRRV 720  
DB 661 KSAFGSSYPAVFVGTIGVGTGAYRSDDCGTTWVLINDDQHOYGNWQQAITGDHANLRRV 720  
QY 721 YIGTNGRGIVYDGGAPSG 740  
DB 721 YIGTNGRGIVYDGGAPSG 740

;; TITLE OF INVENTION: CELLULOYTICUS  
;; FILE REFERENCE: 40197.4US01  
;; CURRENT APPLICATION NUMBER: US/09/917,376  
;; CURRENT FILING DATE: 2001-07-28  
;; NUMBER OF SEQ ID NOS: 7  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 1  
;; LENGTH: 957  
;; TYPE: PRT  
;; ORGANISM: Acidothermus cellulolyticus  
;; FEATURE:  
;; NAME/KEY: MOD RES  
;; LOCATION: (957)  
;; OTHER INFORMATION: Any amino acid  
US-09-917-376-1

Query Match 100.0%; Score 4036; DB 12; Length 957;  
Best Local Similarity 100.0%; Pred. No. 3.2e-301;  
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ATTQPYTWSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW	60
Db	47	ATTQPYTWSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW	106
QY	61	NNWNGVGVSIADPINTNKWAAVGMVYNSWDNDGAILRSSDQATWQITPLPKLGG	120
Db	107	NNWNGVGVSIADPINTNKWAAVGMVYNSWDNDGAILRSSDQATWQITPLPKLGG	166
QY	121	NMPGRGMRERLAVDPNNNDILYFGAPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDT	180
Db	167	NMPGRGMRERLAVDPNNNDILYFGAPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDT	226
QY	181	TCYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNNPVFWSRDGGATWQAVPGAPTGF	240
Db	227	TCYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNNPVFWSRDGGATWQAVPGAPTGF	286
QY	241	IPHKGVFPVNVHLVIATNTGCPYDGGSDGVKFSVTSQGTWTRISPVSTDTANDYFGY	300
Db	287	IPHKGVFPVNVHLVIATNTGCPYDGGSDGVKFSVTSQGTWTRISPVSTDTANDYFGY	346
QY	301	SLGTIDRQHPNTIMVATQISWMPDTIIIFRSTDGGATWTRIDWTSYPNRSRLRYLDISAE	360
Db	347	SLGTIDRQHPNTIMVATQISWMPDTIIIFRSTDGGATWTRIDWTSYPNRSRLRYLDISAE	406
QY	361	PWLTFGVQPNPVPSPKLGWDEAMAIDPFNSDRMLYGTGATLYATNDLTWKWDSGGQIHI	420
Db	407	PWLTFGVQPNPVPSPKLGWDEAMAIDPFNSDRMLYGTGATLYATNDLTWKWDSGGQIHI	466
QY	421	APMVKGLBETAANDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSPVFTTGTSDY	480
Db	467	APMVKGLBETAANDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSPVFTTGTSDY	526
QY	481	AELNPSIIVRAGSPDPSSQPNDRHVAFSTDGKWNFGQSEPGGVTTGGTVAASADGSRFV	540
Db	527	AELNPSIIVRAGSPDPSSQPNDRHVAFSTDGKWNFGQSEPGGVTTGGTVAASADGSRFV	586
QY	541	WAPGDPGPVVYAVGVFGNSWAASQGVPAQAQIRSDRVNPKTFYALNSGTFFRSTDCGVTF	600
Db	587	WAPGDPGPVVYAVGVFGNSWAASQGVPAQAQIRSDRVNPKTFYALNSGTFFRSTDCGVTF	646
QY	601	QPVAAGLPSSGAVGVNMFHVPKEGDLWLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG	660
Db	647	QPVAAGLPSSGAVGVNMFHVPKEGDLWLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG	706
QY	661	KSAPGSSYPVAVVGTIGCVTGAYRSDCCGTWVLINDDQHQYGNWGAITGDHANLRV	720
Db	707	KSAPGSSYPVAVVGTIGCVTGAYRSDCCGTWVLINDDQHQYGNWGAITGDHANLRV	766
QY	721	YIGTNGRGIVYDGIIGGAPSG 740	
Db	767	YIGTNGRGIVYDGIIGGAPSG 786	

RESULT 6  
US-10-155-400-1  
;; Sequence 1, Application US/10155400  
;; Publication No. US20030108988A1  
;; GENERAL INFORMATION:  
;; APPLICANT: DING SHI-YOU  
;; APPLICANT: ADNEY, WILLIAM S.  
;; APPLICANT: VINZANT, TODD B.  
;; APPLICANT: HIMMEL, MICHAEL E.  
;; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS  
;; TITLE OF INVENTION: CELLULOYTICUS  
;; FILE REFERENCE: NREL 01-36A  
;; CURRENT APPLICATION NUMBER: US/10/155,400  
;; CURRENT FILING DATE: 2002-10-22  
;; NUMBER OF SEQ ID NOS: 7  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 1  
;; LENGTH: 957  
;; TYPE: PRT  
;; ORGANISM: Acidothermus cellulolyticus  
;; FEATURE:  
;; NAME/KEY: MOD RES  
;; LOCATION: (957)  
;; OTHER INFORMATION: Any amino acid  
US-10-155-400-1

Query Match 100.0%; Score 4036; DB 14; Length 957;  
Best Local Similarity 100.0%; Pred. No. 3.2e-301;  
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ATTQPYTWSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW	60
Db	47	ATTQPYTWSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW	106
QY	61	NNWNGVGVSIADPINTNKWAAVGMVYNSWDNDGAILRSSDQATWQITPLPKLGG	120
Db	107	NNWNGVGVSIADPINTNKWAAVGMVYNSWDNDGAILRSSDQATWQITPLPKLGG	166
QY	121	NMPGRGMRERLAVDPNNNDILYFGAPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDT	180
Db	167	NMPGRGMRERLAVDPNNNDILYFGAPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDT	226
QY	181	TCYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNNPVFWSRDGGATWQAVPGAPTGF	240
Db	227	TCYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNNPVFWSRDGGATWQAVPGAPTGF	286
QY	241	IPHKGVFPVNVHLVIATNTGCPYDGGSDGVKFSVTSQGTWTRISPVSTDTANDYFGY	300
Db	287	IPHKGVFPVNVHLVIATNTGCPYDGGSDGVKFSVTSQGTWTRISPVSTDTANDYFGY	346
QY	301	SLGTIDRQHPNTIMVATQISWMPDTIIIFRSTDGGATWTRIDWTSYPNRSRLRYLDISAE	360
Db	347	SLGTIDRQHPNTIMVATQISWMPDTIIIFRSTDGGATWTRIDWTSYPNRSRLRYLDISAE	406
QY	361	PWLTFGVQPNPVPSPKLGWDEAMAIDPFNSDRMLYGTGATLYATNDLTWKWDSGGQIHI	420
Db	407	PWLTFGVQPNPVPSPKLGWDEAMAIDPFNSDRMLYGTGATLYATNDLTWKWDSGGQIHI	466
QY	421	APMVKGLBETAANDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSPVFTTGTSDY	480
Db	467	APMVKGLBETAANDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSPVFTTGTSDY	526
QY	481	AELNPSIIVRAGSPDPSSQPNDRHVAFSTDGKWNFGQSEPGGVTTGGTVAASADGSRFV	540
Db	527	AELNPSIIVRAGSPDPSSQPNDRHVAFSTDGKWNFGQSEPGGVTTGGTVAASADGSRFV	586
QY	541	WAPGDPGPVVYAVGVFGNSWAASQGVPAQAQIRSDRVNPKTFYALNSGTFFRSTDCGVTF	600
Db	587	WAPGDPGPVVYAVGVFGNSWAASQGVPAQAQIRSDRVNPKTFYALNSGTFFRSTDCGVTF	646
QY	601	QPVAAGLPSSGAVGVNMFHVPKEGDLWLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG	660
Db	647	QPVAAGLPSSGAVGVNMFHVPKEGDLWLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG	706

[illegible]

Db 352 DISNAPWLODITSTDOFPV---RVGMMVEALALDPFDSNHMLYGTGLTVYGGHDLTNWDS 408  
QY 415 GQIHIAPMVKGLEETAVNDLISPPSGAPLISALGDLGGFTHADVTAVPSTFTSPVFTT 474  
Db 409 KHNVTVKSALVAGIEEMAVLGLITPPGPPALLSAGVDDGGFYHSDLDAAAPNQAHTPTYGT 468  
QY 475 GTSVDYAEINPSIIYRAGSFDPSOPDRHVAFSTDGKNWFOGSEPGCVTTGGTVAASA 534  
Db 469 TNGIDYAGNKPNSIYRSGASD--DYPT---LALSSNFGSTWYADYAASTSTGTGAVALSA 523  
QY 535 DGSRRVWAPGDPQPVVAVGFGNSWAASQGVPAANAQIRSDRVNPKTFYALNSGTFYRST 594  
Db 524 DGDVTLLMSSSTGALVSQSG---TLTAVSSLPSCAVIASDKSDNTVYFGSAGAIYVSK 580  
QY 595 DGVVTFQPVAAAGLPSSGAVGVMFHAPVKGEDLWLAASGLYHSTNGGSSWSAI-TGVSS 653  
Db 581 NTATSFTKTVS-LGSSTTVNAI-RAHPSIAGDVMASTDKGLWHSTDYGSTFTQIGSGVTA 638  
QY 654 AVNVGFGKSAPGSSYPVAVFVGTIGVTCAYRSDDCGTTWVLINDDQHOYGNWQA--IT 711  
Db 639 GWSFGFGKASSTGSYVVIYGFPTIDGAAGLFKSEDAAGTNQWVISDASHGFGS-GSANVNV 697  
QY 712 GDHANLRVYIGTNGRGIVYGDIGAPSG 740  
Db 698 GDLQTYGRVFRGHERPHLLRSQREPAG 726  
RESULT 9  
US-10-155-400-7  
; Sequence 7, Application US/10155400  
; Publication No. US20030108988A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS  
; FILE OF INVENTION: CELLULOXYLICUS  
; CURRENT APPLICATION NUMBER: US/10/155, 400  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 7  
; LENGTH: 726  
; TYPE: PRT  
; ORGANISM: Aspergillus aculeatus  
US-10-155-400-7

Query Match 41.6%; Score 1680; DB 14; Length 726;  
Best Local Similarity 46.2%; Pred. No. 2.7e-120;  
Matches 346; Conservative 113; Mismatches 258; Indels 32; Gaps 17;  
QY 1 ATTQPTWNSVNAI-GGGGFDGIVNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVG 59  
Db 1 AASQAYTWKVVTVGGGGGFTPGIVNPSAKGVAYARTDIGGAYRLN-SDDTWTPLMDWVG 59  
QY 60 ---NNWNYGVVSTAADPINTNKVAAVGMVYNSWDPNDGAILRSSDQAGTQITPLPF 116  
Db 60 NDTWHDW---GIDALATPEVDTRVYVAVGMVYNEWDNPVGSILRSTDQGTWTETKLPP 116  
QY 117 KLGNNMPGRGGERLAVDPNNINILYFGAPSGKGLWRSTDSGATWSQMTNFPDVGTIYAN 176  
Db 117 KVGNNMPGRGGERLAVDPNNKNSILYFGARSGHGLWKTSDYGATWSNVTSTWTGTYQD 176  
QY 177 PTDFTGYQSDIQGVVWVAFDSSSSSLGQASKTIFVGVADPNPNVFWSDGGATWQAVPGA 236  
Db 177 SSST--YTSDPVGIAMVTFDSTSGSSGSATPRIFVGVADAGKSVFKSEDAAGTAAWVSGE 234  
QY 237 PT-GFIPHKGFDPVNVHLYIATSTGPGYDSSGDVWKFVSTGTTWTRISFVPSSTDTAN 295  
Db 235 PQYGFPLHKGLVSPEKTLTYISYANGAGPYDGTNGTHVKNYITSGVMTDISP---TSLAS 291

QY 296 DYFGVSGLTIDRQHPNTIMVATQISWMPDITIIIPRSTDGGATWTRINDWTSYPNRSIRYVL 355  
Db 292 TYGYGGLSVDLQVPGTLMVAALNCWMPDELIPRSTDGATWSPIEWWNGYPSINTYYSY 351  
QY 356 DISASBPWLTFFGVQPNP-PVPSPKLGWMDRAMAIDPNSDRMLYGTGATLVATNDLTKWDS 414  
Db 352 DISNAPWLODITSTDOFPV---RVGMMVEALALDPFDSNHMLYGTGLTVYGGHDLTNWDS 408  
QY 415 GQIHIAPMVKGLEETAVNDLISPPSGAPLISALGDLGGFTHADVTAVPSTFTSPVFTT 474  
Db 409 KHNVTVKSALVAGIEEMAVLGLITPPGPPALLSAGVDDGGFYHSDLDAAAPNQAHTPTYGT 468  
QY 475 GTSVDYAEINPSIIYRAGSFDPSOPDRHVAFSTDGKNWFOGSEPGCVTTGGTVAASA 534  
Db 469 TNGIDYAGNKPNSIYRSGASD--DYPT---LALSSNFGSTWYADYAASTSTGTGAVALSA 523  
QY 535 DGSRRVWAPGDPQPVVAVGFGNSWAASQGVPAANAQIRSDRVNPKTFYALNSGTFYRST 594  
Db 524 DGDVTLLMSSSTGALVSQSG---TLTAVSSLPSCAVIASDKSDNTVYFGSAGAIYVSK 580  
QY 595 DGVVTFQPVAAAGLPSSGAVGVMFHAPVKGEDLWLAASGLYHSTNGGSSWSAI-TGVSS 653  
Db 581 NTATSFTKTVS-LGSSTTVNAI-RAHPSIAGDVMASTDKGLWHSTDYGSTFTQIGSGVTA 638  
QY 654 AVNVGFGKSAPGSSYPVAVFVGTIGVTCAYRSDDCGTTWVLINDDQHOYGNWQA--IT 711  
Db 639 GWSFGFGKASSTGSYVVIYGFPTIDGAAGLFKSEDAAGTNQWVISDASHGFGS-GSANVNV 697  
QY 712 GDHANLRVYIGTNGRGIVYGDIGAPSG 740  
Db 698 GDLQTYGRVFRGHERPHLLRSQREPAG 726  
RESULT 10  
US-10-420-191-2  
; Sequence 2, Application US/10420191  
; Publication No. US200400067569A1  
; GENERAL INFORMATION:  
; APPLICANT: No. US20040067569A1ozymes Biotech, Inc.  
; APPLICANT: Rey, Michael W.  
; APPLICANT: Zaretsky, Elizabeth J.  
; APPLICANT: Haas, Jeffrey A.  
; TITLE OF INVENTION: Polypeptides Having Xyloglucanase Activity And Nucleic Acids  
; FILE OF INVENTION: Encoding Same  
; FILE REFERENCE: 10210-200-US  
; CURRENT APPLICATION NUMBER: US/10/420,191  
; CURRENT FILING DATE: 2003-04-18  
; PRIOR APPLICATION NUMBER: US 60/373,987  
; PRIOR FILING DATE: 2002-04-19  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: Patent In version 3.2  
; SEQ ID NO 2  
; LENGTH: 838  
; TYPE: PRT  
; ORGANISM: Trichoderma reesei  
US-10-420-191-2

Query Match 39.1%; Score 1579; DB 12; Length 838;  
Best Local Similarity 43.0%; Pred. No. 1.9e-112;  
Matches 325; Conservative 123; Mismatches 264; Indels 44; Gaps 20;  
QY 6 YTWSNVAI-GGGGFDGIVNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWV---GW 60  
Db 21 FSWKVKVLGGGGGFPVGIIFHPKTKGVAYARTDIGGLYRLN-ADDSWTAVTDGIADNAGW 79  
QY 61 NNWNYGVVSTAADPINTNKVAAVGMVYNSWDPNDGAILRSSDQAGTQITPLPFKLG 120  
Db 80 HNW---GIDAVALDQDDQKVVAAGVYNSWDPNSGAILRSSDRCATWSFTNLPPKVG 136  
QY 121 NMPGRGGERLAVDPNNINILYFGAPSGKGLWRSTDSGATWSQMTNFPDVGTIYANPTD 180  
Db 137 NMPGRGGERLAVDPANSNIIYFGARSGNGLWKTGGTGGVTSKVSFTATGTYPDPSDS 196



Qy 181 TGYOSDIQGVVWVAFKSSSLGQASKTIFVGVADP--NNPVFWRDGGATWQAVPGAPTG 239  
Db 197 NGYNSDKQGLMWVTFDSTSTTGCATGRIFVGTADNITASVYVSTNAGSTWSAIVPGPGK 256  
Qy 240 FIPHKGVDPVNVHLYIATNTGTPYDGGSSGVKFSVTSCTWTRISPVSTDTANDYFG 299  
Db 257 YFPHAKLOPAEKALYITSDGTGYPDGLGVSVMRYDIAGTGWKDIITPVSGSDL---YFG 313  
Qy 300 YSGLTIDRQHNTIMVATQISWNPDTIIFRSTGGATWTRIDWTSYPNRSLRYVLDISA 359  
Db 314 FGLGLDLQKPGTLVVASLNSWPDQALFRSTDGTTWSPINAWASYPTETYYYSISTPK 373  
Qy 360 EPWLTFG--VQPNPPVPS-----PKLGWDEAMAIIDPFNSDRMLYGTGATLYATNDLTKWDS 414  
Db 374 APWIKNNFIDVTSBSPSDGLIKRLGWMIESLEIDPTDSNHWLYGTGWTIFGCHDLTNWDT 433  
Qy 415 GGQIHAPMVKGLEETAIVNDLISPPSGAPLISALDGLGFTHA---DVTAVPSTIFTSPV 471  
Db 434 RHNVSIQSLADGIEEFVSQDLASAPGSELLAAVGDNDGFTFASRNDLGTSPQTVWATPT 493  
Qy 472 FTTGTSVDYAEINPSIIVRAGSPDPSPQPNDRHVAFTDGGKNWFOGSEPGGVTTGGTVA 531  
Db 494 WATSTVDYAGNSVKSVVRVGN-----VAISDGGATWSDIYAADTSMNGGTVA 548  
Qy 532 ASADGSRFVWAPGDPGPVYVAVFGNSWAASQGVPAANAQIRSDRVNPKTFYALSNGTFY 591  
Db 549 YSADGDTILWSTASSG---VQRSQFGSFAVSLLPAGAVIASDKKTNVSFYAGSGSTFY 605  
Qy 592 RSTGGVTFPQVAAGLPSSGAVGM--FHAVPGKEGDLWLAASGLYHSTNGGSSWSAI- 648  
Db 606 VSKDTGSSF---TRG-PKLGASGTTIRIDAAHPTTAGTLYVSTDVGVIFRSTDSGTTFGQVS 661  
Qy 649 TGVSSAVNVGFKSAPGSSYPVAVFVGTIGGVTGA--YRSDCGTTWVLINDDQHQYGNW 706  
Db 662 TALNTYQIALGVGS-GSNW-NLYAFGT--GPSGARLYASGDSGASWTDIQGSGFGSID 717  
Qy 707 QQAITGDHANLRVYVGTNGRGIVY--GDIGGAPSG 740  
Db 718 STKVAGSGSTAGQVYVGTNGRGVFAOQTVGGGTGG 753

RESULT 12  
US-10-026-994-2  
; Sequence 2, Application US/10026994  
; Publication No. US20030113732A1  
; GENERAL INFORMATION:  
; APPLICANT: Dunn-Coleman, Nigel  
; APPLICANT: Goedegebuur, Frits  
; APPLICANT: Ward, Michael  
; APPLICANT: Yao, Jian  
; TITLE OF INVENTION: EGVI Endoglucanase and Nucleic Acids  
; TITLE OF INVENTION: Encoding the Same  
; FILE REFERENCE: GC698  
; CURRENT APPLICATION NUMBER: US/10/026,994  
; CURRENT FILING DATE: 2002-04-30  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 818  
; TYPE: PRT  
; ORGANISM: Trichoderma reesei  
US-10-026-994-2

Query Match 39.0%; Score 1573.5; DB 14; Length 818;  
Best Local Similarity 42.9%; Pred. No. 4.9e-112;  
Matches 324; Conservative 122; Mismatches 265; Indels 45; Gaps 20;

Qy 6 YTNWNAI-GGGGFVDGIVNEGAPGLYVTRDTIGMYRWDAANGRWIPLLDWV---GW 60  
Db 2 FSKNVKVLGGGGVPGIIFPHKTKGVAYARTDGGYRLN-ADDSWATVTDGIADNAGW 60  
Qy 61 NNWYNGVVSIAADPINTNKVAAVGMVYNSWDNDGAILRSSDQGWATQITPLPFKLG 120

Db 61 HNW---GIDAVALDDQDKVVAVGMVTNSWDPNSGAILRSSDRGATWSFTNLPRKVG 117  
Qy 121 NMPGRGEMERLAVDPNNDNIIYFGAPSGKGLWRSTDSGATWQMTNFPDVGTYIANPTDT 180  
Db 118 NMPGRGAGERLAVDPANSNIIYFGARSGNGLWKSTDGGVTFKVSSTATGTVIPDPSDS 177  
Qy 181 TGYOSDIQGVVWVAFKSSSLGQASKTIFVGVADP--NNPVFWRDGGATWQAVPGAPTG 239  
Db 178 NGYNSDKQGLMWVTFDSTSTTGCATGRIFVGTADNITASVYVSTNAGSTWSAIVPGPGK 237  
Qy 240 FIPHKGVDPVNVHLYIATNTGTPYDGGSSGVKFSVTSCTWTRISPVSTDTANDYFG 299  
Db 238 YFPHAKLOPAEKALYITSDGTGYPDGLGVSVMRYDIAGTGWKDIITPVSGSDL---YFG 294  
Qy 300 YSGLTIDRQHNTIMVATQISWNPDTIIFRSTGGATWTRIDWTSYPNRSLRYVLDISA 359  
Db 295 FGLGLDLQKPGTLVVASLNSWPDQALFRSTDGTTWSPINAWASYPTETYYYSISTPK 354  
Qy 360 EPWLTFG--VQPNPPVPS-----PKLGWDEAMAIIDPFNSDRMLYGTGATLYATNDLTKWDS 414  
Db 355 APWIKNNFIDVTSBSPSDGLIKRLGWMIESLEIDPTDSNHWLYGTGWTIFGCHDLTNWDT 414  
Qy 415 GGQIHAPMVKGLEETAIVNDLISPPSGAPLISALDGLGFTHA---DVTAVPSTIFTSPV 471  
Db 415 RHNVSIQSLADGIEEFVSQDLASAPGSELLAAVGDNDGFTFASRNDLGTSPQTVWATPT 474  
Qy 472 FTTGTSVDYAEINPSIIVRAGSPDPSPQPNDRHVAFTDGGKNWFOGSEPGGVTTGGTVA 531  
Db 475 WATSTVDYAGNSVKSVVRVGN-----TAGTQVAISSDGGATWSDIYAADTSMNGGTVA 528  
Qy 532 ASADGSRFVWAPGDPGPVYVAVFGNSWAASQGVPAANAQIRSDRVNPKTFYALSNGTFY 591  
Db 529 YSADGDTILWSTASSG---VQRSQFGSFAVSLLPAGAVIASDKKTNVSFYAGSGSTFY 585  
Qy 592 RSTGGVTFPQVAAGLPSSGAVGM--FHAVPGKEGDLWLAASGLYHSTNGGSSWSAI- 648  
Db 586 VSKDTGSSF---TRG-PKLGASGTTIRIDAAHPTTAGTLYVSTDVGVIFRSTDSGTTFGQVS 641  
Qy 649 TGVSSAVNVGFKSAPGSSYPVAVFVGTIGGVTGA--YRSDCGTTWVLINDDQHQYGNW 706  
Db 642 TALNTYQIALGVGS-GSNW-NLYAFGT--GPSGARLYASGDSGASWTDIQGSGFGSID 697  
Qy 707 QQAITGDHANLRVYVGTNGRGIVY--GDIGGAPSG 740  
Db 698 STKVAGSGSTAGQVYVGTNGRGVFAOQTVGGGTGG 733

US-10-156-761-10111  
; Sequence 10111, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 10111  
; LENGTH: 739  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-10111

```
Query Match      34.8%; Score 1405; DB 14; Length 739;
Best Local Similarity 40.8%; Pred. No. 3.8e-99;
Matches 308; Conservative 104; Mismatches 257; Indels 86; Gaps 21;

QY 6 YTSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAANGRWIPLLDWGWNNWGY 65
DB 42 YRNRNAVIGGTFVTVGLVPHFVSVRGLAYARTDIGGAYRWDGGRARWTLIDHLGWDNNL 101
QY 66 NGVSVIAADPINTNKVMAAVGMYTNSWDPNDGAILRSSDOGATWQITLPEKLGNNMGR 125
DB 102 LGVEMAVDPHPLDYLAVGTYAQSWAGN-GAVLRSEDRGATWTRTDLTVKLGNEGR 160
QY 126 GNGERLAVDPNNILYFPGAPSGKGLWRSTDSGATWSQMTNFPDVGTYIANPTDTTGYQ 185
DB 161 GAGERLLVDPDSDTLWGT-RHDGLLKSTDRGATAAAATAPP-----AKANS 207
QY 186 DIQGVVWVAFPKSSSLGQASKTIFVGVADPNP-----VWSDRGGATWQAVCAPGPF 240
DB 208 SGQGVWF-----LVAAGRTVYAGWDGDSGTGTANLYTAD-GTTWGAVPGRPSGT 257
QY 241 --IPHKGVFDPVNVHLYIATSNCGPYDGGSDVWKFVSQVTSWTRISPV-----PSTDT 293
DB 258 SAKVPLRAAYTHRELIVTVGDAPCGQSDGSHKLRATGATWTEVTPVKPGTTSDG 317
QY 294 ANDYFGYGLTIDRQHPNTIMVATQISWMPDTIIFRSTGGATWTRIWDWTSYPNRSRY 353
DB 318 SADTFAYGGVAVDARRPCTLVVSTNNRWADGTVFRSTGGRTWTSKX-----AA 368
QY 354 VLDLSAPFWLTFGVQNPVPSPKLGWDEAMADPPNSDRMLYGTGATLYATNDLTW 413
DB 369 VFDVSETPFLDWGDD-----KPKFGMWIQALAVDPYDSQHVYGTGATLYTRDLKEW- 421
QY 414 SGGQIHIAAPMKGLEETAUNDLISPPSG-APLISALDGLGFTHADVTAVPST-IFTSPV 471
DB 422 -----APRIRGLEESAVROLISPPVGEAHLISGLDIGVYHERLTASPSRGATNPV 474
QY 472 FTTGTSVDYAEINPSIIVRAGSFDPSSQPNDRHVAFASTDDGKNW--FGSEPPGVVTTGGT 529
DB 475 FGSATGLAQAAARPAVYVTRTGWGDHNG-----AYSHDGGRTWAPFEAQPIAKDAPG 528
QY 530 VAASADGSRFWA-----PGDQGPVVYA-----VGFNGSWAASQGVPAQAQIRSRVNPKT 581
DB 529 IATSDAGGTLTWSFVHWDG-----TTYAHRSTONGASWSVSPKPGATPVADPADPTR 583
QY 582 FYA--LNGTFYRSTDCGVTQFPAAGLPSSGAVGMFHAVPGEGDLWLAAS-SGLYHS 638
DB 584 FYADFNGTLTASTDSGRSFTARAGGLP-SGDSQFKLVAAPGRSGDLWLSAKWNGLYRS 642
QY 639 TNGGSSSAITGVSSAVNVGFKSAPSSYPVAVFVVGVTIGGVTGAYRSDDCGTTWVLND 698
DB 643 TDGGDTFARIDSCWASYTLFGKAADGADYPAIQVGSFTEITAVYRSDDAARTWVRIND 702
QY 699 DOHQYGNWQAITGDHANLRVYIGTNGRGIVYGD 733
DB 703 DAHQWGWIGEAVGDPRIHGRVYLATNGRGIQYGE 737

RESULT 13
US-10-395-241-14
; Sequence 14, Application US/10395241
; Publication No. US20040038367A1
; GENERAL INFORMATION:
; APPLICANT: YAOI, Katsuro
; TITLE OF INVENTION: NOVEL XYLOGUCAN OLIGOSACCHARIDE-DEGRADING ENZYME, POLYNUCLEOTIDE
; FILE REFERENCE: Q73756
; CURRENT APPLICATION NUMBER: US/10/395,241
; PRIOR APPLICATION NUMBER: JP 2002-83433
; PRIORITY FILING DATE: 2003-03-25
; NUMBER OF SEQ ID NOS: 19
```

```
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 14
; LENGTH: 789
; TYPE: PRT
; ORGANISM: Geotrichum sp. M128
; US-10-395-241-14

Query Match      27.7%; Score 1117; DB 12; Length 789;
Best Local Similarity 35.0%; Pred. No. 5.8e-77;
Matches 281; Conservative 124; Mismatches 279; Indels 120; Gaps 31;

QY 6 YTSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAANGRWIPLLDWGWNNWGY 65
DB 4 YEFKNVAIGGGYITGIVAHPKTKDILYARTDIGGAYRWDAAGTSKWIPLNDFIEAQDMNI 63
QY 66 NGVSVIAADPINTNKVMAAVGMYT-NSWDPNDGAILRSSDOGATWQITLPEKLGNNMGP 124
DB 64 MGTESIALDPNDPDRLLYLAQRYVGDW-----AAFYVSEDRGQSFITYESFPFMGANDMG 119
QY 125 RGMGERLAVDPNNILYFPGAPSGKGLWRSTDSGATWSQMTNFPDVGTYIANPTDTTGYQ 184
DB 120 RNNGERLAVNPNSNEVWMTGT-EGIKSSDRAKTWTNVTSPDAF-----TNGIGYT 172
QY 185 SDIQGVVWVAFPKSSSLGQASKTIFVGVADPNPVPFWSRDGGATWQAVPGATGFI--- 241
DB 173 S-----VIFDP-----ERNGTIYASATAPQG-MYVTHDGGVSEVPVAGQPSSSLNRT 218
QY 242 -----PHKGVFDPVNVHLYIATSNCGPYDGGSDVWKFVSQVTSWTRISPV----- 285
DB 219 TGAFDPKPKASTAPQPMKVALTP--NFLVTVYADYVPGMGTTFGEVNRQNRITSGAMDDIT 276
QY 286 -----SPVSTDTANDYFGYGLTIDRQHPNTIMVATQISWMPDTI---IFRSTDCGAT 336
DB 277 PRVGNSSPAPYNNQTFPAGGFCGLSDVATNPENLVIT-LDRDPGALDLSIYLDAGAT 335
QY 337 WTRIWDWTS-----YPNRSRLRYVLDISAEPMWITFGVQP-----NPVPSPG---KLG 379
DB 336 WKDVTQLSSPSNLEGNGWGHPTNAARY-KDGTVPVPLDFFNNGPQWGGYGAPHGTPGLTKFG 394
QY 380 WDEAMADPPNSDRMLYGTGATLYATNDLTK-----WDSGGQIHIAAP---MVKGLEETA 431
DB 395 WMSAVLIDPFPPEHLMYGTGATWATDLSRVEKDW-----APSWYLQIDGIEENA 446
QY 432 VNDLISPPSGADPLISALDGLGFTHADVTAVPSTIFTSPVTTGTSVDYAEINPSIIVRA 491
DB 447 ILSLSRPSKGAALLIGIGDI-SMKHDDLTK-FQKMFGAQFNSLDSIDAAGNFPNVVRA 505
QY 492 GSFDPSQPNDRHVAFASTDDGKNW--FGSEPPGVVTT---GGTVAASADGSRFWAP--G 544
DB 506 GSSGHEYDSACARGAVATDGDATWTFPTCPPGMNASHYQGSTIAVDASGSIWSTKLD 565
QY 545 DPGQPVVYAVGRGNSWAASQGVPA-----NAQIRSDRVNPKTFYALSNGTFFRSTDCGV 598
DB 566 EQASGFWYSHDYKGTWS-----VPAGDLKAQTANVLSDKVQDGTFFYATDGGKFFVSTDDGK 621
QY 599 TPOPVAAGLPSSGAVGMFHAVPGEGDLWL-AAASGLYHSTNGGSSWSAI-TGVSSAVN 656
DB 622 SYAAKGAGLVTT--GTSLMPAVNPWAGDVWVPEGGLFHSTDFGASFTVGTGANATLVS 679
QY 657 VGFGKS-----APGSSYPVAVFVVGTT--IGGVTGAYRSDDCGTTWVLINDDOHQYGNWQAI 710
DB 680 VGAPKSKSGKASAPSAPVFIWTDKPGSDIGLYRSDNGSTWTRVNDQEHNTSG-PTMI 738
QY 711 TGDHANLRVYIGTNGRGIVYGD 734
DB 739 EADPKVYGRVYLTNGRGIVYADL 762

RESULT 14
US-10-395-241-12
; Sequence 12, Application US/10395241
; Publication No. US20040038367A1
; GENERAL INFORMATION:
; APPLICANT: YAOI, Katsuro
; TITLE OF INVENTION: NOVEL XYLOGUCAN OLIGOSACCHARIDE-DEGRADING ENZYME, POLYNUCLEOTIDE
; FILE REFERENCE: Q73756
; CURRENT APPLICATION NUMBER: US/10/395,241
; PRIOR APPLICATION NUMBER: JP 2002-83433
; PRIORITY FILING DATE: 2003-03-25
; NUMBER OF SEQ ID NOS: 19
```



Qy	545	DPGPPVYAVGFCNSWAASQVPA-----NAQIRSDRVNPKTFYALSNGTFRSTDCGV	598
Db	567	EQASGPMYSHDYGKTWS-----VPAGDLKAQTANVLSKXVDGTFYATDGGKFPVSTDCGK	622
Qy	599	TFQPVAAGLPSSGAVGVNMFHAVPGKEGDLWL-AASSGLYHSTNGGSSWSAI-TGVSSAVN	656
Db	623	SYAAKAGLVT--GTSIMPVNPVWAGDVWVPVPEGGLFHSTDFGASFTRVGTANATLVS	680
Qy	657	VGFGKS----APGSSYPAVFVVGTT--IGGVTCAYRSDDCGTTWVLINDDDHQYGNWGOAI	710
Db	681	VGAPKSKDGKKASAPSAVFIMGTDKPGSDIGLYRSDDNGSTWTRVNDQEHYSG-PTMI	739
Qy	711	TGDHANLRRVYIGTNGRGIVYDI	734
Db	740	EADPKVYGRVYLTNGRGIVYADL	763

Search completed: May 11, 2004, 12:25:47  
Job time : 90.3468 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 12, 2004, 03:40:05 ; Search time 10636.8 Seconds  
(without alignments)  
3015.368 Million cell updates/sec

Title: US-09-917-376-3

Perfect score: 4036

Sequence: 1 ATTQPTWNSVAIGGGGVD.....YIGTNGRGIYVGDIGGAPSG 740

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DRV=xlh  
-Q=/cgm2\_1/USPTO.spool/US09917376/runat\_11052004\_114535\_28241/app\_query.fasta\_1.1429  
-DB=GenEmbl -OFMT=fastp -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -GFMT=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORW=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09917376 @CGN 1 1 7742 @runat\_11052004\_114535\_28241 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:\*  
1: gb.ba:\*  
2: gb.htg:\*  
3: gb.in:\*  
4: gb.om:\*  
5: gb.ov:\*  
6: gb.pat:\*  
7: gb.ph:\*  
8: gb.pl:\*  
9: gb.pr:\*  
10: gb.ro:\*  
11: gb.sts:\*  
12: gb.sy:\*  
13: gb.un:\*  
14: gb.vi:\*  
15: em.ba:\*  
16: em.fun:\*  
17: em.hum:\*  
18: em.in:\*  
19: em.mu:\*  
20: em.om:\*  
21: em.or:\*  
22: em.ov:\*  
23: em.pat:\*  
24: em.ph:\*  
25: em.pl:\*  
26: em.ro:\*  
27: em.sts:\*  
28: em.un:\*

RESULT 1

ALIGNMENTS

Result No.	Score	Query Match	Length	DB ID	Description
1	4036	100.0	2869	6	AX700058 Sequence
2	2429.5	60.2	299800	1	AP005028 Streptomy
3	2371	58.7	296500	1	SC0939128 Streptomy
4	2063	51.1	2823	6	AX565635 Sequence
5	2061.5	51.1	3262	1	AF078038 Caldicell
6	2009	49.8	12732	1	AE007608 Clostridi
7	1982	49.1	2950	1	AJ585344 Clostridi
8	1753.5	43.4	2579	8	AB1292929 Agarius
9	1722	42.7	2745	8	AB015511 Aspergill
10	1620	40.1	3959	8	AY040839 Aspergill
11	1579	39.1	2517	8	AY281371 Hypocrea
12	1560.5	38.7	2724	8	AK110506 Oriza sat
13	1405	34.8	299550	1	AP005031 Streptomy
14	1162.5	28.8	14520	1	AE011809 Xanthomon
15	1160.5	28.8	13278	1	AE001712 Thermotog
16	1145	28.4	11910	1	AE012276 Xanthomon
17	1117	27.7	2367	6	AX924431 Sequence
18	1117	27.7	2646	6	AX924429 Sequence
19	1117	27.7	2716	8	AB089343 Geotrichu
20	1114	27.6	2481	6	AX924435 Sequence
21	460.5	11.4	590	8	AB1534359 Agarius
22	368.5	9.1	24208	1	AY281358 Unculture
23	271.5	6.7	47537	7	AY328853 Vibrio pa
24	268.5	6.7	301950	1	AP006570 Gloeobact
25	262	6.5	58128	7	AY357582 Burkholde
26	261	6.5	49575	7	AY328852 Vibrio pa
27	238	5.9	20342	1	AE004613 Pseudomon
28	226.5	5.6	300134	1	AE017211 Geobacter
29	218.5	5.4	7938	1	AE011079 Methanosa
30	214	5.3	5600	1	AF182518 Pseudomon
31	214	5.3	303226	1	AE016774 Pseudomon
32	206	5.1	34866	1	BX640426 Bordetell
33	204.5	5.1	306050	1	RME603645 Rhizobium
34	198.5	4.9	349497	1	BX640440 Bordetell
35	193.5	4.8	348257	1	BX640425 Bordetell
36	193	4.8	303400	1	AP006578 Gloeobact
37	192.5	4.8	207212	2	AL671876 Mus muscu
38	192	4.8	346362	1	BX640439 Bordetell
39	192	4.8	342416	1	BX842573 Mycobacte
40	189	4.7	150894	1	D90904 Synecocyst
41	186.5	4.6	130001	1	D90915 Synecocyst
42	186	4.6	212199	9	AL445199 Human DNA
43	186	4.6	306650	1	AP005221 Corynebac
44	185.5	4.6	2454	4	AF178428 Bos tauru
45	185.5	4.6	4161	1	BSF133614 Bacillus

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

ALIGNMENTS

AX700058  
LOCUS AX700058 2869 bp DNA linear PAT 03-APR-2003  
DEFINITION Sequence 2 from Patent WO03012090.  
ACCESSION AX700058  
VERSION AX700058.1 GI:29536021  
KEYWORDS  
SOURCE  
ORGANISM  
Acidothermus cellulolyticus  
Acidothermus cellulolyticus  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Frankineae; Acidothermaceae; Acidothermus.  
1.  
REFERENCE  
Ding, S. Y., Adney, W. S., Vinzant, T. B. and Himmel, M. E.  
Thermal tolerant avicelase from Acidothermus cellulolyticus  
Patent: WO 03012090-A 2 13-FEB-2003;  
Midwest Research Institute (US)  
FEATURES  
Location/Qualifiers  
1..2869  
/organism="Acidothermus cellulolyticus"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:28049"  
ORIGIN  
Alignment Scores:  
Pred. No.: 2,48e-198 Length: 2869  
Score: 4036.00 Matches: 740  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0  
US-09-917-376-3 (1-740) x AX700058 (1-2869)  
Qy 1 AlaThrThrGlnProTyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAsp 20  
Db 139 GCGACGACTCAGCGGTACACTCGGAGCAACGTGGCGATCGGGCGCGCTTGTTCGAC 198  
Qy 21 GlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGly 40  
Db 199 GGGATCGTCTTCAATGAAGGTGCACCGGGAATCTGTACGTGGGACGGACATCGGGGG 258  
Qy 41 MetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrp 60  
Db 259 ATGTATGATGGATGGATGCGGCAACGGCGGTGGATCCCTCTTCTGGATGGGTGGATGG 318  
Qy 61 AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 80  
Db 319 AACAAATGGGGGTACACGGCGTCTGACGATTCGGGACACCGGATCAATACATCAAG 378  
Qy 81 ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 100  
Db 379 GTATGGCGCGCGTGGGAATGTACACCAACAGCTGGGACCCAAACAGACGGAGGATTC 438  
Qy 101 ArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGly 120  
Db 439 CGTCTGCTGATCAGGCGCAACGTGGCAATAACGCCCTCGCTTCAAGCTTGGCGGC 498  
Qy 121 AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle 140  
Db 499 AACATGCCCGGGCGTGGAAATGGGAGCGGCTTCGGGTGGATCCAAACATATGACAACT 558  
Qy 141 LeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThr 160  
Db 559 CTGTATTTCGGCGCCCGGAGCGGCAAGGGCTCTGGAGAGACACAGATTCCGGCGCGACC 618  
Qy 161 TrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAlaProThrAspThr 180  
Db 619 TGGTCCACAGATGACAACTTTCGGGACGTAGGACGCTACATTGCCAATCCCACTGACACG 678  
Qy 181 ThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSer 200  
Db 679 ACCGGCTATCAGAGCGATATTCAGGGCGTCTGGGTCTGCTTCGACAAAGTCTTCGTCA 738  
Qy 201 SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProVal 220

Db 739 TCCTCGCGCAAGCGAGTAAAGACCAATTTTGTGGCGGTGGCGGATCCCAATTAATCCGGTTC 798  
Qy 221 PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyValaProThrGlyPhe 240  
Db 799 TTCTGGAGACAGACGGCGCGGCGAGTGGCGGTCGCCGGTGGCGGACCGGCTTC 858  
Qy 241 IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsn 260  
Db 859 ATCCCGCACAGGGCGTCTTTGACCCGCTAACACAGCTGCTCTATATTGCCACACCAAT 918  
Qy 261 ThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGly 280  
Db 919 ACGGGTGGTCCGTATGACGGGAGCTCCGGCGACGCTCGGAAATTCGGTGACCTCCGGG 978  
Qy 281 ThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyr 300  
Db 979 ACATGGACGCGAATCAGCCCGTACCTTCGACGGACACGGCCACAGCACTACTTGGTAC 1038  
Qy 301 SerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSer 320  
Db 1039 AGCGGCTCACTATCGACCGCGACGACCCGCAACAGATAATGGTGGCAACCCAGATATCG 1098  
Qy 321 TrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIle 340  
Db 1099 TGGTGGCGGACACCAATTAATCTTTCGGAGCACCGCGCGGTGGCGGACCGGATC 1158  
Qy 341 TrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGlu 360  
Db 1159 TGGGATGGACGAGTATCCCAATCGAAGCTTGCATATGTCTTGACATTTCCGGCGGAG 1218  
Qy 361 ProTrpLeuThrPheGlyValGlnProAsnProValProSerProLysLeuGlyTrp 380  
Db 1219 CCTTGGCTGACCTTTCGGGTACAGCCGAATCCTCCGTTACCCAGTCCGAAGCTCGGTCG 1278  
Qy 381 MetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGly 400  
Db 1279 ATGGATGAAGCGATGGCAATCGATCGTCAACTCTGTATCGGATCTCTACCGAACAGSC 1338  
Qy 401 AlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIle 420  
Db 1339 GCGAGCTTGTACGCAACAAATGATCTCAGAAAGTGGAGCTCCGGCGGCGCAGATTCATATC 1398  
Qy 421 AlaProMetValLysGlyLeuGluGluThrAlaValAsnAspLeuIleSerProSer 440  
Db 1399 GCGCCGATGGTCAAGGATTCGAGGAGACGGCGGTAAACGATCTCATCAGCCCGCGTCT 1458  
Qy 441 GlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThr 460  
Db 1459 GGCGCCCGCTCATCAGCGCTCTCGGAGACCTCGCGGCTTCCACCACCGCGCAGCTTACT 1518  
Qy 461 AlaValProSerThrIlePheThrSerProValPheThrGlyThrSerValAspTyr 480  
Db 1519 GCGGTGCATCGACGATCTTCACGTTCACCGTGTTCACGACGGCAGCGGCTCGACTAT 1578  
Qy 481 AlaGluLeuAsnProSerIleValArgAlaGlySerPheAspProSerSerGlnPro 500  
Db 1579 GCGGAATTGAATCGTCCGATCATCGTTCGCGCTGGAGTTTCGATCCATCCAGCAACCG 1638  
Qy 501 AsnAspArgHisValAlaPheSerThrAspGlyLysAsnTrpPheGlnGlySerGlu 520  
Db 1639 AACGACAGGACGCTCGCGTCTTCGACACAGCGCGCAAGAACTGGTTCACAGCGACGCA 1698  
Qy 521 ProGlyGlyValThrThrGlyThrValAlaAlaSerAlaAspGlySerArgPheVal 540  
Db 1699 CCTGGCGGGTGAACGACGGCGGACCGCTCGCGCATCGCGCGACGGCTCTCTGTTTCGTC 1758  
Qy 541 TrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrp 560  
Db 1759 TGGGCTCCGGCGATCCCGGTACGCTGTGTGTACGAGTCCGGATTTGGCAACTCTCTCG 1818  
Qy 561 AlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLys 580

(E-mail: bio@nite.go.jp, URL: <http://www.bio.nite.go.jp/>,  
Tel: 81-3-3481-1933, Fax: 81-3-3481-8424)

This work was done in collaboration with Haruo Ikeda (\*1), Jun  
Ishikawa (\*2), Akiharu Hanamoto (\*3), Chigusa Takahashi (\*3), Mayumi  
Shinose (\*3), Hiroshi Horikawa (\*4), Hidekazu Nakazawa (\*4), Tomomi  
Osonoe (\*4), Norihiro Kuchida (\*4), Hiseashi Kikuchi (\*4), Tadayoshi  
Shiba (\*5), Yoshiyuki Sakaki (\*6, \*7), Masahira Hattori (\*1, \*7)  
and Satoshi Omura (\*1, \*3).

Final finishing process and all annotation were done by H. Ikeda  
and J. Ishikawa.

- \*1 Kitasato Institute for Life Sciences, Kitasato University
- \*2 National Institute of Infectious Diseases
- \*3 The Kitasato Institute
- \*4 National Institute of Technology and Evaluation
- \*5 School of Science, Kitasato University
- \*6 Institute of Medical Science, University of Tokyo
- \*7 RIKEN, Genomic Sciences Center

Following url is also available.  
<http://avermitilis.is.kitasato-u.ac.jp>.

Location/Qualifiers

```
1. -299800
/organism="Streptomyces avermitilis MA-4680"
/mol_type="genomic DNA"
/db_xref="taxon:227892"
/note="this strain is also named as strain: ATCC 31267,
NCIMB 12804 or NRRL 8165.-synonym: Streptomyces
avermectinius"
16. 918
/note="SAV1722"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
```

```

/db_xref="GI:2960367"
/translation="WAGRALVSCADLTLPMWTSABSDGPRVDRLLRYPVKSMLGDE
LPVETBERGLAHDRAVALDRDTGVASAKHPLMSRLLTVTQAQDNSAVRMTSDGG
KELSDTEGSDBALSEIVGRPVFLTDAPHEAALDRADVEDLURDHTRPVPSARTGR
GGSGEGEFTDFPAPVHVMTSTLARI AALAPNGSAQARYRNVILRPLTRGFTEND
WDTRLHLGTGLVLVRIARTPRCAITPLAHGTVPADTAALRVLAENHRVTPTESMGPO
PCAGVYQVLRPGRIRLGGPPVHA"
complement(1232..3406)
/note="SAV1723"
/codon_start=1
/transl_table=11
/product="putative mycodextranase"
/protein_id="BAC69434.1"
/db_xref="GI:29603369"
/translation="MHSNLSRYVRMSATGAVALAAGLLVLAAPRAAAGADLPTFT
SYEASATTTGKIGDFTQGTIAVSEASQAVRLGAGQVEFTFPRANAANVSVYNN
PQSGSALNVYNGTKLAKTAVTSKYVSVDTSWIAKATTHPFDNRALLGNQVQPG
DKIAEVASTVTDVADVEQVAAATKAGAVSVTDKGADPSQGDSTQAFRAIAAAS
AQQVWVPPGQYRLTSSINGVYVTLQAGSGWHSVVHTSRFTDQSSSGVHLKDFAP
VIGETERVDSNPFVNGSLRGSSVSGMWLQHLKVGLWTGNNDLVVNNRILDT
TADGLNLNAGRVNRNNFLRNGDDALAMSLYSPDTNSSFESNTISQPNLANGLA
IYGGTDLAKNNLISDTNAGLSIAKSNQGLDLPSPLAGTITVSGNTLVVRAGAMNP
WQHPVGLRVDSYDGAIEANVSTGTTITDPSYSAFVSGGGHGATKNTVTDGATV
RNTGTVVAQESQGAKEFNVOATGGAIGVNCYPAGSGTFTLTDGGGAGWSSTW
GDCSAPQCGNPDPPTENLAKGPATATGSDQVYTPCKAVDGDASTYWESNNFAF
PQATVLDGSSQAVRRLKLPATNAQKATQTLVSQGSTDCSAYSTVWASQGYRFD
TGNATVTLPSGTNNLYLKLNVNTANTGMPAAQFSEVAYLS"
complement(3464..5065)
/gene="aglA1"
/complement(3464..5065)
/gene="eglA1"
/note="SAV1724"
/codon_start=1
/transl_table=11
/product="putative alpha-glucosidase"
/protein_id="BAC69435.1"
/db_xref="GI:29605369"
/translation="MGQPTPARIQKWHTVATVYQVYVRSFADGDGDTGDLGAVRAB
LPYLAGLGVDAIWFNFPNGSKGGYDVADYRAVDPAFGTLEAKLIIAEARELGR

```

```

TLVDIVPNHVSQHPWFRALAAAGCPSPERALFHFGRGCAHGEPLPNDWPSQFAGST
EPWTLQPDGMDTLHLTFDEQPDNDWHAFPAQROEHEDVLRFPWFERGVAGVYRISAAALL
AKDPLDVLGAPRHPFDEQDELHDIYRSWRAIADEYGAVFGEVWMLPDPTEFARYL
RDELTAFTAFSFLSCFWDARRLRITAIDETAEHAFVGPATWLCNHDVTRTVTRYG
RTDGTDFAAKAGFTPTDLAALGRRARAAMALLSLALPGSVLYLGBELGLPEAEIIPRE
SIQDPMHFRSHGVDRPCDRCRVLPLWADAPHAGFGEPWLPQAPASWAVAADLOAED
PSMSLYLEALRLPAYADGLPTWLPAPGVULAFARTEGLMCMVNLAUTPAGLPAHS
RULLSSGPDQDQRLPQDTAAWLRI"
/notes="SAV1725"
/notes="SAV1725"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="BAC69436.1"
/db_xref="GI:29605370"
/translation="MSTRILVSPILSRPRGKAYTVFTGVVLLFALAFLPVYVWV
TGAMKSPDEARTPPSLVPQKWLHSGVTDADLMQLPAHLNWTVOATGAWLLQLVFC
TAAAYALSLKVPFGLVGLLGLATLMVPAQALVVPKYLTVADLP.LHLSLLNDPLGI
WLPVANAFLYLLKRFDDQIPRDVLEAAIDGGRRLTWSIVLPMRSRPLVGVSLF
ALVAVQDPLFWLPMVFSDDTKQPIVALVQLSQNIQLTVLIAAMVIAASIPMWAMFLVF
QRHIIAGISAGSTKG"
complement (5969..6925)
/notes="SAV1726"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="BAC69437.1"
/db_xref="GI:29605371"
/translation="MTKTAARRPVEAIVPVPAPPPGGGLGRRLADQAHAFLLG
GLLCFALFSWPYAIRAVIAFOKYTGSPPEWGTANTFRVLHDPEFTAAWNLITFT
LUALLIGFAVFFVWVNLHAKAFRRVYVWVLPVMI PPVVSALLWKVFPDPAAGLA
NEALRLFTSNWNGATDALSVIVATWANNGGTVLIIAALQGIPOGLYEAEAL
DGANILQRHFTPIPTOTRFTVLMQLLIITATMQVTEPFVITGGGPENATVTVLYLI
KYAFIYDFGGACALSVMLLVLLGAFSAVYLRLLTSABEDS"
complement (6931..8271)
/notes="SAV1727"
/codon_start=1
/transl_table=11
/product="putative sugar-binding lipoprotein precursor"
/protein_id="BAC69438.1"
/db_xref="GI:29605372"
/translation="MSSSGFRACVALVACSFVLTCGSSGDAADGKTRITVNCMPPK
SAKVDKRFEDDIATFKNPDIDVVAHDAFPDQDPKTFDAKLQAGMEDVFTYFD
ATHVDVNOQAADITSVKOLKSIDYVQQLRDIYTVDGKYGVPRTGYSMLGILYNNFL
FKSADGKDPKPTDEVAAAKRIAAAGGTGVYADYSAQNOGQWHFTAEYLSQGD
VYADGKATVDTPPEKAVIQLNLDHMTWDTNMGSKQLIIVNDVQOLMGSGKLMGLYS
APDNIPILVKSGTDTDLALAMPCKGTLIGDGMFKKATPAQIKAGLKLWLFPM
FLTPGKGFIDYARAKNDAPVGLPPEPLRFLSGAADARDQVKKANANVPVENTYQFEL
GNQSRMKIPEPQAOQIYSVLDSSAVSLTKDADIDKLLKDSGKIDSILARG"
complement (8418..9410)
/notes="SAV1728"
/codon_start=1
/transl_table=11
/product="putative LacI-family transcriptional regulator"
/protein_id="BAC69439.1"
/db_xref="GI:29605373"
/translation="MTERRAEVAKKXVSEATVSRVLNDKPGVSPATROAVLSALDVL
GYERPOLGERARLVGLVPELQNPFPAPAEVIGGALAQLGLTVLCTQTKGGVSE
ADYVNLQQQVSGVFGAGLYAQADPHDHRRLADRNIPVVLVNAALIEHLGFPVVS
CDDAVAEQWRHLASHGIERIGLVGPDHMPFSARKLAARALGDLDPDEYVARAFS
IEGHAASRLIDRGVGTICASDPLALGAVRAARKGLDVPQISIVVGYDDSLAMNC
TEPPLTVPQPTESMGRAVVELLNARIGSSVPDELLPELVLVVRGSTAQAPR"
complement (9457..10185)
/notes="SAV1729"
/codon_start=1
/transl_table=11
/product="putative GntR-family transcriptional regulator"
/protein_id="BAC69440.1"
/db_xref="GI:29605374"
/translation="MSGDPAPVPGDPAPRKRRTAAQHSAPLTLGGRKESRYERVAD
ALRAALIAELRAGEVYSAPLAAFGVVSATPVREAMLDLIAKEGLVDVTPVKNKGRVTA
VSEKQJDEYTHRSLEIETPTAGLTADPAELEALRPVAREIYVAATAGDLIAYVEA
DLRFLGLLALAGNDHLVDVDRRLRRSRRLYGLTALARAGRLRASAEHELLELDALIA

```

```

RDTFAAREVTRHLGHVRLWVAP"
complement (10198..11199)
/notes="SAV1730"
/codon_start=1
/transl_table=11
/product="putative proline racemase"
/protein_id="BAC69441.1"
/db_xref="GI:29605375"
/translation="MRSKLVHLHVDHSHTGMPTRVITGGIGTIPGATMNERLYFHEH
RDDIROLIMNEPRGHSGAVLOPPARPDCDVGVIYEVSGYLPSCGHGTIGVATVL
VETGMVEVPEVTTRIDTTPAGLVAVAEVDEGAARAVTLKNVPSFCVGLDRKATLAD
GRVTYDLAYCGNEFYAILPLGEFGLPDRSRKODILKAGLSMDAVNADGGVHPEDP
SIRCHVHLVAPGSDARHSAHHPGWFDRSPCGTGSARMAQLHARGELPLGTE
FVNESLITRTFTGRLGTTEVAGIPAVLPSPFTGRAWVTGTATQYLLDPPDFPAGFVL"
complement (11214..12104)
/notes="SAV1731"
complement (11214..12104)

Alignment Scores:
Pred. No.: 1.25e-113 Length: 299800
Score: 2429.50 Matches: 438
Percent Similarity: 73.99% Conservative: 111
Best Local Similarity: 59.03% Mismatches: 174
Query Match: 60.20% Indels: 19
DB: 1 Gaps: 8

US-09-917-376-3 (1-740) x AP005028 (1-299800)

Qy 6 TyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn 25
Db 175758 TACAGCTGGAAGAAGCCGCCGCTCGACCGCGCGCTTCGCCCCGGCATCGTCTTCAAC 175817
Qy 26 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrpAsp 45
Db 175818 CGCTCCGAGAGAACCCTCGCTAGCCCGCACCGACATCGCGCGCGCTACCGCTGGGCC 175877
Qy 46 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr 65
Db 175878 GAGTCTCGAAGACCTCGACGCGCTGCTGACCTCGCTCGGCTCGGAGCGACTGGGGGCAC 175937
Qy 66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 85
Db 175938 ACGGCTGTCGAGACCTCGCTCGCATCCGTCGACCCGCAACAAAGGTGTACCGCGCGCTC 175997
Qy 86 GlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGln 105
Db 175998 GGCACGTACACGAACAGCTGGGACCCGGGCAACGCTGCGCTCTCAGGTCCGCGCACCGG 176057
Qy 106 GlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArg 125
Db 176058 GGCAGAGCTGGCAGAACAGCCGCTGCTGCTTCAAGCTGGGCGGGAACATCGCGGCGCG 176117
Qy 126 GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAla 145
Db 176118 GGCATGGCGGAGCGCTCGCGCTCGACCCGCAACAGAACAGCGTGTGTATTCGCGCGG 176177
Qy 146 ProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThr 165
Db 176178 CCCAGCGCAAGGGCTGTGGCGTTCAGGACTCGGGGCGCTCTGTCGCGAGGTACCC 176237
Qy 166 AsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSer 185
Db 176238 GACTTCCGACGCTCGGCACCTACGTGCGAGCGGACCGCACACGAGCGGGTACGCGTCC 176297
Qy 186 AspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerSerLeuGlyGlnAla 205
Db 176298 GACAACGAGGCATCGTGTGGGTCACTTCAGCAGTGCAGCGGTGCGCGGGAGCTCC 176357
Qy 206 SerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSerArgAsp 225
Db 176358 ACGCGGACGGTACGTGGGTGCGGCGTCCGCCACAGCAACTCCGTCTATCGCTCCACCGAC 176417
Qy 226 GlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIleProHisLysGly 245

```



Db 176418 GCGGCGGCGAGCTGGTCCCGGCTGGCGGCGAGCGCCAGCGGCATCTCGGCCCAACAGGCG 176477  
 Qy 246 ValPheAspProValAenHisValLeuTyrIleAlaThrSerAenThrGlyGlyProTyr 265  
 Db 176478 GTGCTGACGCGGCGAGCGCTGCTGTACTCGCTGACGAGCAGGCGGAGCGGACCTGAC 176537  
 Qy 266 AspGlySerSerGlyAspValTyrLysPheSerValThrSerGlyThrTrpThrArgile 285  
 Db 176538 GACGGCGGAGGAGCAGCTGTGGCGGTACACGACGAGAACCGCGGACCTGGACGAACATC 176597  
 Qy 286 SerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIle 305  
 Db 176598 AGCCCGCTCGCGAGGCGGACAC-----TACTACGGCTTCAGCGGCTGACCGGTG 176648  
 Qy 306 AspArgGlnHisProAenThrIleMetValAlaThrGlnIleSerTrpTrpProAspThr 325  
 Db 176649 GACCGGAGCATCCGGGAGCGGTGATGGCGACTCGGTACAGCTCTGGTGGCGGAGCAGC 176708  
 Qy 326 IleIlePheArgSerThrAspGlyGlyValAlaThrTrpThrArgIleTrpAspTrpThrSer 345  
 Db 176709 CAGCTCTTCGCTCCAGGACGCGGCGGCACTGGAGGAGGCTGGGACTACACCTCG 176768  
 Qy 346 TyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpThrPhe 365  
 Db 176769 TATCCGAGCGCTCGAACCGCTTCACCATGGATGCTCGCTCGCTCGCTCGCTCGCTCGCT 176828  
 Qy 366 GlyValGlnProAsnProValProSerProLysLeuGlyTrpMetAspGluAlaMet 385  
 Db 176829 GGAGCGAACCCCGGACCGCGGAGCAGACCCGGAACCTCGGCTGGAGTCCGCTG 176888  
 Qy 386 AlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAla 405  
 Db 176889 GAGATGACCCGCTCGACTCGCGCGCATGATGACGAGACGGCGGCGGCGGCTACGCGC 176948  
 Qy 406 ThrAsnAspLeuThrIleTyrAspSerGlyGlyGlnIleHisIleAlaProMetValLys 425  
 Db 176949 ACGGACAACTGACGAACCTGGGACAGCGGAGGAGCGAGTTCACCATCAAGCCGATGGCGCG 177008  
 Qy 426 GlyLeuGluThrAlaValAsnAspLeuIleSerProSer---GlyAlaProLeu 444  
 Db 177009 GGCTTGAGAGAGAGCGCGCTCAACAGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCG 177068  
 Qy 445 IleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaValProSer 464  
 Db 177069 TTACGCGCTCGGTGATCGCGGCTTCGGGACAGCGACTTCACCGAGTTCGCGGCTCG 177128  
 Qy 465 ThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeuAen 484  
 Db 177129 CTGATGTACAGTCCGCGAATCTTACACGAGCAGCAGCTCGACTACGCGGAGCGGAC 177188  
 Qy 485 ProSerIleIleValAlaGlySerPheAspProSerSerGlnProAsnAspArgHis 504  
 Db 177189 CCGGCGAGGTGTGGCGGTTCGGAATCTCGAC-----TCGGGTTCG-----CAT 177233  
 Qy 505 ValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGlyVal 524  
 Db 177234 GTGCGTTCGACGAGAACCGCGGCAATGTTTCGGGGGCGGAGCCCTTCGGGGGTC 177293  
 Qy 525 ThrThrGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaProGly 544  
 Db 177294 AGCGGGGTGGAGCGTTCGGCGGCGGCTCGAGCGGAGTCTGCTGTTGGAGCGCGG 177353  
 Qy 545 AspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAspSerGln 564  
 Db 177354 GGCACCGGG-----GTGACGTACACGACCGGGTTCGGCACCTCGTGTGGCGGCTCGCG 177407  
 Qy 565 GlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrAla 584  
 Db 177408 GGCTTCGCGGCGGCGGATCGCTGAGTCCGAGCGGTTCGACCGGAGACTCTTCTACGCG 177467  
 Qy 585 LeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProValAla 604

Db 177468 TTCAGTCCGCGAGGTTCCTACGTCTGAGTTCGGAGCGGGGCGACCTTACGCGGTCCGG 177527  
 Qy 605 Ala---GlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaValProGlyLys 623  
 Db 177528 GCCACGGGCTTCCGAGCGGCGACAGC-----GTGGCTTCAAGGCGCTGCCGCGCAGC 177581  
 Qy 624 GluGlyAspLeuTrpLeuAlaAspSerSer-----GlyLeuTyrHisSer 638  
 Db 177582 AAGGCGACATCTGCTCGCGGCGGCGAGCAGCGCGCTACGGGTGTGGCACTCG 177641  
 Qy 639 ThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGly 658  
 Db 177642 ACGACGCGCGCGGCGCTTCCACCAAGCTCGCACCGCTGACAGCGCTTACACCGCGGAGATCGGC 177701  
 Qy 659 PheGlyLysSerAlaProGlySerSerTyrProAlaValPheValValGlyThrIleGly 678  
 Db 177702 TTCGGCAAGGCGGCGGCGGCGCTTCTACAGACGCTTACACCGCGGAGATCGGC 177761  
 Qy 679 GlyValThrGlyAlaTyrArgSerAspCysGlyThrTrpValLeuIleAsnAsp 698  
 Db 177762 GGTGTGCGCGCATCTTCGCTCGACCCGACGAGGCGGAGCTGGACCCGCTCAACGAC 177821  
 Qy 699 AspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArg 718  
 Db 177822 GATCCACCACTAGTGGGTGGACGCGCGCGGATCACCGGTGACCCCGGCTCTACGCG 177881  
 Qy 719 ArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyAlaPro 738  
 Db 177882 CGGTGTATGTGTGACGAGCGGCGGAGTCTACGCGGACACCGCGGCTCTTCG 177941  
 Qy 739 SerGly 740  
 Db 177942 GACGCG 177947

RESULT 3  
 SCO939128/c 296500 bp DNA linear BCT 11-FEB-2003  
 LOCUS Streptomyces coelicolor A3(2) complete genome; segment 25/29.  
 DEFINITION AL939128 AL0323862 AL031013 AL031107 AL031350 AL031371 AL031515  
 ACCESSION AL033505 AL034433 AL035161 AL079355 AL096825 AL645882  
 VERSION AL939128.1 GI:24413907  
 KEYWORDS  
 SOURCE Streptomyces coelicolor A3(2)  
 ORGANISM Streptomyces coelicolor A3(2)  
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE 1  
 AUTHORS Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C.H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neill, S., Rabinowitz, E., Rajandream, M.A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietz, R., Woodward, J., Barrell, B.G., Parkhill, J. and Hopwood, D.A.  
 Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2)  
 JOURNAL Nature 417 (6885), 141-147 (2002)  
 MEDLINE 21996410  
 PUBMED 12000953  
 REFERENCE 2 (bases 1 to 296500)  
 AUTHORS Bentley, S.D.  
 DIRECT SUBMISSION  
 TITLE Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces sequencing team, Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: sdb@sanger.ac.uk  
 JOURNAL On or before Oct 26 2002 this sequence version replaced  
 COMMENT gi:20520750, gi:20520810, gi:20520753, gi:20520666, gi:20520757, gi:20520758, gi:20520668, gi:20520759, gi:20520814, gi:20520765, gi:20520767.

FEATURES  
 Location/Qualifiers  
 source 1. .296500

```

/organism="Streptomyces coelicolor A3(2)"
/mol_type="genomic DNA"
/strain="A3(2)"
/db_xref="taxon:100226"
complement(144..443)
/gene="SCO6480"
/notes="synonym: SC9C7.16c"
complement(144..443)
/gene="SCO6480"
/notes="SC9C7.16c, unknown, len: 99 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC9C7.16c"
/protein_id="CAA22728.1"
/db_xref="GI:4154075"
/db_xref="SPTREMBL:Q9ZBJ4"
/translation="MYGCTASMKTRPGFREEVUULLSADGLRAGCELYVVGLE
DDEITLWVTRWTRDASLEPAKDAIGRAMPLTGEFTKQMSVAGGLGL"
complement(517..1563)
/gene="SCO6481"
/notes="synonym: SC9C7.17c"
complement(517..1563)
/gene="SCO6481"
/notes="SC9C7.17c, conserved hypothetical protein, len:
348aa; similar to several hypothetical proteins eg.
SM:Y0BN MYCTU from Mycobacterium tuberculosis (340 aa)
fasta scores; opt: 998, z-score: 1321.8, E(): 0, (47.6%
identity in 355 aa overlap)"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAA22729.1"
/db_xref="GI:4154076"
/db_xref="SPTREMBL:Q9ZBJ3"
/translation="MRWENLTDSDHARAADTALFGADVTRFTDFEPRGIFHEVR
ARSLNRVPGASRMPFETWVYRGCTHACVCFARKTHSYLDDTGLGDFDSQIVVKV
NAPELVRQRQASRQEWGHEVMDVCYQRAEGRYLMFGIIIEALTRANPISILTK
GTLILRDLLTLRAARVTEGIVSVSGFTDAELWRTVEGTPAPERLEVRVRLGGRG
IGCVLMAPVPIPLGDEPAOLRATVRAIAAAGATVTPVLHLRPGAREFWFMAWLGRH
HPLVRRELYADGAYAPKWYQRRITRQVHDLAREGIGIPARAGMFRVAEPHSRP
AEFAPPEPVQLSLI"
1800..2243
/gene="SCO6482"
/notes="synonym: SC9C7.18"
1800..2243
/gene="SCO6482"
/notes="SC9C7.18, conserved hypothetical protein, len: 348
aa; similar to two hypothetical proteins e.g. TR:069685
(EMBL:AL022121) Mycobacterium tuberculosis RV03718c or
MTW025.066c, 147 aa; fasta scores; opt: 475, z-score:
768.3, E(): 0, (52.1% identity in 146 aa overlap)"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAA22730.1"
/db_xref="GI:4154077"
/db_xref="SPTREMBL:Q9ZBJ2"
/translation="MAQVEATERRIVADAETVPDLADYSGTREKLLPHEPSEYVR
EGDGGEGTLVHWKLQATSKRVDCLEVEFTDGLVKKDRNSMVTWTRTPAGECK
SRVVLVTWQAGVGCGGFFETKPAKGLARIYDALLAKLAAEVEK"
2394..4841
/gene="SCO6483"
/notes="synonym: SC9C7.19"
2394..4841
/gene="SCO6483"
/notes="SC9C7.19, probable efflux protein, len: 815aa;
similar to many, several of which are involved in toxin
antibiotic efflux eg. TR:069070 (EMBL:AF061335)
oxytetracycline exporter from Streptomyces rimosus (536
aa) fasta scores; opt: 1020, z-score: 879.9, E(): 0,
(36.9% identity in 542 aa overlap)"
/codon_start=1
/transl_table=11

```

```

/product="putative efflux protein"
/protein_id="CAA22731.1"
/db_xref="GI:4154078"
/db_xref="GOA:Q9ZBJ1"
/db_xref="SPTREMBL:Q9ZBJ1"
/translation="MGGITLVQDPVRDPQDEPVVAAAPGPPVPGDGERDAGLSP
RRVRLVFGMLALLAEQMIIVATAPKIVGELGDKRMSWALTALLTSTVGIPFI
YKGLDIFGRKGVQFAILVFVVGSLAGATQMDQIAFRAVQAGGLMGVQVQAI
IADIVPRERARYMGLICASGLASVAGPLGGLGYFTDLSWRWCIFYINVFGLVTVAV
VAVVLLKLPVRPRKLDITGALLIAAASCTCLVLTSMGCTEYAWGSRVITGLAAAGA
ATTLFLVAEHPAPEPLIPLFRDJSIENVATLVGLVGVALFGAASLYLPTFLQMDVGA
SATESGLMLLPMGCVGASIIISQLISRTGCHYRIHPILGSALSVGMWLLSRLDADT
SRHYSIWMVIGLIGLVMPVLVAVQNSRPTDGLTATANNYFRQIIGSVGAARF
GTLFAGRLTDALADRIIPAEAGVGLPDAEAITPQLVHSLPPALRDYIIGAYDMPRIIF
LYLVPLVLIGLLILFLKEKPLSHNAPVTDPTAPAPQPPAQVPPARSPYAGGI
PVCTGVOHDPCTGVVPRAALTILIDVGGQOIGRSGASDDGRYALATPGSGAYVLIAAGG
HQPQAVSVTVGVRPVELDVGLGAGRLAGSVMTADGSPVRDATVTLTVHVGVRVATTR
SGRGGGVITELVAGETILASAPAFRAALPVSQAAARETRQDVVELAGGAVLRTVR
ASGRVADARVTLDDAAGNVVDLTITGPDGTFRFRVLDSSGEYTVIAAGYPVATVQLQ
VAGGRTERDLQLGHED"
4911..7493
/gene="SCO6484"
/notes="synonym: SC9C7.20"
4911..7493
/gene="SCO6484"
/notes="SC9C7.20, conserved hypothetical protein, len:
860aa; similar to many hypothetical proteins eg.
TR:086875 (EMBL:U22894) from Streptomyces lividans (547
aa) fasta scores; opt: 1000, z-score: 738.2, E(): 0,
(40.0% identity in 497 aa overlap) and TR:086807
(EMBL:AL031031) from Streptomyces coelicolor (916 aa)
fasta scores; opt: 358, z-score: 316.9, E(): 2.3e-10,
(27.2% identity in 779 aa overlap). Also contains a short
region of similarity to TR:Q53897 (EMBL:X60316) AbaA
regulatory locus from Streptomyces coelicolor (192 aa)
fasta scores; opt: 108, z-score: 223.2, E(): 3.8e-05,
(38.5% identity in 135 aa overlap)".
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAA22732.1"
/db_xref="GI:4154079"
/db_xref="GOA:Q9ZBJ0"
/db_xref="SPTREMBL:Q9ZBJ0"
/translation="MVTGHPRTVVTATQILRSRGQRPDAMDHGTGAPPQHRAGPD
TVPADPVTGRVPLAVVDRDGLYSHWSRGARLFLALKEAIGEPADLLPVSGLVP
EDADPDGAGHDDSTRSTADHDPGPGFLESSLDGRISYPAAGRARTIVRADRDVLMW
AYPLVPGPERLLVLAADADGLRDYADGTDSDAGPAVVERIAPAFALHTDPPGDEL
ARRLEILPMSVGESARIYAQVLELGYVPLEFSQHDVPVPTDGMVPRRAERRE
RAARLAAAGEPVPDLDEADLELYAAVRELEFLNEVSGAIGTSLDLSRTIVEVSRA
VVPRTDVAGTYLREQVVGAGFPGDVPDITTMHVRVALEHTDEPRDWDVVVGEAM
PPAHTFPFCMTTGEPVLPRIEEMGHAIASQEKDIREPLTGRSMVLPLKART
VVLGFMILLRHPERPVFNDMDRVTGAEALAGLVLDNARMYTHQENVALTQDSMLP
HIPARLACDIATRYLPTGLRGVGMFADVKLPFGARTLVVGMVGHGLNSAAMWG
QLTRVAVQTMAGLIDLPQALLRNLDLAQRLDTHLATCLYAYDPIAGELHLANAGHI
PPVLVRAEDGASLIDLTPTAGTIGVGVVPFESVTVVAPDRLVMTDGLVEVRGEDI
GVGLATCESAAHPAASMDACDTIIRALGAGGGRKODVALLMARLTGTEPDAAVEWR
LARDPVEGRRAAVREQLYDWGICPLVAPRELMSVELTVNAVRHSHARPVALRLVRA
DTLLCEVDVDDDDHPELPALKSAGPDETGRLRVSTLAKEWGASRTGAGKTVMFELTP
ARRH"
7711..8433
/gene="SCO6485"
/notes="synonym: SC9C7.21"
7711..8433
/gene="SCO6485"
/notes="SC9C7.21, unknown, len: 240 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC9C7.21"
/protein_id="CAA22733.1"
/db_xref="GI:4154080"
/db_xref="GOA:Q9ZBJ9"
/db_xref="SPTREMBL:Q9ZBJ9"

```

/translation="MSVTSRYRDWEGFNEASGDRGAVFWDAEPTLTAGRHILAHFPP  
 HLTDAALPLIDLGGNGTQTRYFAERPHVVDGADLSAALAHARHADPPGLAAYRQLD  
 AADKAQALHAEGLANVTVRGVILHQCPEDDROPLADLAALAHARHADPPGLAAYRQLD  
 AKTVLMGLAQGPAGPPAKLAPLVRHLGLTPGAVPDEAVPEYLRAAGLTVALSGELPLVT  
 TXYTAGTRELPSAWVAGTA"  
 8603. .9436  
 /gene="SC06486"  
 /notes="synonym: SC9C7.22"  
 8603. .9436  
 /gene="SC06486"  
 /note="SC9C7.22, possible transport associated protein,  
 len: 277aa; similar to SW:DPFA BACSU dppa, dipeptide  
 transport protein (274 aa) fasta scores; opt: 448,  
 z-score: 604.3, E(): 2.3e-26, (32.3% identity in 269 aa  
 overlap)."  
 /codon\_start=1  
 /transl\_table=11

## Alignment Scores:

Pred. No.:	1.24e-110	Length:	296500
Score:	2371.00	Matches:	431
Percent Similarity:	72.29%	Conservative:	109
Best Local Similarity:	57.70%	Mismatches:	183
Query Match:	58.75%	Indels:	24
DB:	1	Gaps:	10

US-09-917-376-3 (1-740) x SCO939128 (1-296500)

Qy	6	TyrThrTrpSerAsnValAlaIleGlyGlyPheValAspGlyIleValPheAsn	25
Db	70439	TACACCTGGGAAGACGCCGATCGACGGGGCGGTTCGTCCCGGCATCGTCTCAAC	70380
Qy	26	GluclyAlaProGlyIleLeuTyrValArgThrAspIleGlyMetTyrArgTrpAsp	45
Db	70379	CGSACGAGAAAGACCTGGCGCTACGCCCGCACACATCGGAGCGCTACCGCTGGCAG	70320
Qy	46	AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr	65
Db	70319	GAGAGTCCACACCTGGAGCGCGCTCTCGACACATCGTGGCTGGAGACTGGGGCAC	70260
Qy	66	AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal	85
Db	70259	ACGGCGGTGCTCTCGCGCTCGACCGCGCTCGATCGGACCGGTCTACCGCGCGGTC	70200
Qy	86	GlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerAspGln	105
Db	70199	GGCAGTACACCAACGACTGGGACCCGACCAACGCGCGGTGCTCGCGCGCGG	70140
Qy	106	GlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArg	125
Db	70139	GGTGGAGCTGGAGAGGGGACCTCGGTTCAGCTGGGCGGCAACATGCCCGCGCG	70080
Qy	126	GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAla	145
Db	70079	GGCATGGTGGAGCGCTGGCGCTGACCCGACCAACGACGCTGTACCTGGCGGCC	70020
Qy	146	ProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThr	165
Db	70019	CCGACGGCCACGACTGTGGCGGTCCACGACGCGGCGGTACCTGGTCCGAGGTGACG	69960
Qy	166	AsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSer	185
Db	69959	GCCTTCCGAAACCCGGAACTACGCCAGACCCGAAACGACGCTCCGCTACGCTCC	69900
Qy	186	AspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerSerLeuGlyGln	204
Db	69899	GACAACGAGGCATCACTGGGTCACTTCGACGAGTCGAGTCGAGCGCGCGCGGACG	69840
Qy	205	AlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSerArg	224
Db	69839	GCCACGGGACCTCTACGTGGGGTTCGCGACAAAGGAGAACGCGGTCTACCGCTCGAC	69780
Qy	225	AspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIleProHisLys	244

Db	69779	GACCGCGCGGACCTGGGAGCGGTCCGCGGCGACCGCCGCGGTACCTGGGCCCAAG	69720
Qy	245	GlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyPro	264
Db	69719	GGTGTCTGGAGCGGAGAACCGGTACCTGTACTCTCGCTACGACACACCGCGGCGCCG	69660
Qy	265	TyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrpThrArg	284
Db	69659	TACGACGCGGCAAGGCGCGGTGTACCGGTACGCGACGCGGACCGCACCTGGAGGAC	69600
Qy	285	IleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThr	304
Db	69599	ATCAGCCCGCGCGGAGCGGACAC-----TACTACGGCTTCAGCGCTCGAC	69549
Qy	305	IleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrpProAsp	324
Db	69548	GTGACCGCGGAGCGCGGCGGTGTGGGACCGCGGTACAGCTCGTGTGGCGGAC	69489
Qy	325	ThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIleTrpAspTrpThr	344
Db	69488	ACGCAGATCTTCGCTCGACGACGCGGCGGACATGTCCAGCGGTGGAGCTACAC	69429
Qy	345	SerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpLeuThr	364
Db	69428	TCGTACCGCGACCGGAGAACCGGTACACGATGAGCTCTCTCGTCCCTCGCTCACC	69369
Qy	365	PheGlyValGlnProAsnProValProSerProLysLeuGlyTrpMetAspGluAla	384
Db	69368	TGGGCGCGAACC-----CGCGCGCGGACGACACCGCGAAGCTGGGCTGGATGAC	69309
Qy	385	MetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyr	404
Db	69308	CTGGAGATCGACCGCTTCGACTCGGACCGGATGATGACGACGCGGCGGACCGGTCTAC	69249
Qy	405	AlaThrAsnAspLeuThrLysTrp---AspSerGlyGlyGlnIleHisIleAlaProMet	423
Db	69248	GGCACGAGAACCTGACGAACTGGGACGACGAGGCGGCGACGTTCCGCGTCGAGCGGATG	69189
Qy	424	VallysGlyLeuGluThrAlaValAsnAspLeuIleSerProProSerGlyAlaPro	443
Db	69188	GTGCGGGCGCTGGAGGAGACGCGCTCAACGACCTGGCGCTCCCGCGCTCGGTCGCGCG	69129
Qy	444	LeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaValPro	463
Db	69128	CTGCTCAGCGCTGGTGACGTCGCGGCTTCGCGCACACGAGCTGACCGAGGTCCG	69069
Qy	464	SerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeu	483
Db	69068	TCGATGATGATACCTCGCGCAACTTCACCTCGACGACGAGCTGGACTTCGCGGAGACG	69009
Qy	484	AsnProSerIleValArgAlaGlySerPheAspProSerSerSerGlnProAsnAspArg	503
Db	69008	AAGCGGACGCTGTGTGTCGCGGCGCAATCTCGAC-----TCCGCTCG-----	68964
Qy	504	HisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGly	523
Db	68963	CACATCGGTCTTCACGACCAACCGCGCAACTGGTTCGCGGCGACCGACCGCTCGCGG	68904
Qy	524	ValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPro	543
Db	68903	GTGAGCGCGGCGACCGTTCGCGGCGGCGGAGCGGCGGCGCGCTTCGTGTGGAGCGCG	68844
Qy	544	GlyAspProGlyGlnProValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSer	563
Db	68843	-----GAGGCTGCGCGGTGAGTACACGACCGGCTTCGCTACGCTCGGCGGCGGTGCG	68790
Qy	564	GlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyr	583
Db	68789	ACCGCGCTCCCGCGGCGGCGATCGTACGAGTCGAGTCGAGCGGCGGCGGCGGCGGCTTC	68730
Qy	584	AlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProVal	603

Db 68729 GGCTTCAAGTCCGGAGGTCTTACGTACGTACGACGGCGCGGACCTTACCGGCTCG 68670  
Qy 604 AlaAala---GlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaValProGly 622  
Db 68669 GCGCGACCGGGCTGCGCGCGGTGAC-----GGCGTCCGCTTCAAGGCGCTGCGGGC 68616  
Qy 623 LysGlyGlyAspLeuTrpLeuAlaAalaSerSer-----GlyLeuTrpHis 637  
Db 68615 GGGGAGGGCGACGCTCTGGCTGGCGCGCGCGGCGGACCGCGCTGACGGCGCTGCGCAC 68556  
Qy 638 SerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnVal 657  
Db 68555 TCAGCGACCGCGCGGACCTTCCAGCGCTGCGCGCGGCGGACCGTCCGCGGCGGACCGTCC 68496  
Qy 658 GlyPheGlyLysSerAlaProGlySerSerTrpProAlaValPheValGlyThrIle 677  
Db 68495 GGCTTCGGAAGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 68436  
Qy 678 GlyGlyValThrGlyAlaValArgSerAspCysGlyThrThrTrpValLeuIleAsn 697  
Db 68435 GCGCGGCTGCGCGGACCTTCCGCTCCAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 68376  
Qy 698 AspAspGlnHisGlnTrpGlyAsnTrpGlyAlaIleThrGlyAspHisAlaAsnLeu 717  
Db 68375 GACGACCGCCACGAGTGGGCTGACCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 68316  
Qy 718 ArgArgValTrpIleGlyThrAsnGlyArgGlyValGlyValGlyAspIle----- 734  
Db 68315 GCGCGGCTGACGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 68256  
Qy 735 ---GlyGlyAlaProSerGly 740  
Db 68255 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 68235

RESULT 4  
AX565635 2823 bp DNA linear PAT 29-NOV-2002  
LOCUS Sequence 1 from Patent WO02077242.  
DEFINITION AX565635  
ACCESSION AX565635  
VERSION AX565635.1 GI:26000969  
KEYWORDS  
SOURCE unidentified  
ORGANISM unclassified.  
REFERENCE 1  
AUTHORS Duffner, F. and Sjoelholm, C.  
TITLE Family 74 xyloglucanases.  
JOURNAL Patent: WO 02077242-A 1 03-OCT-2002;  
Novozymes A/S (DK)  
FEATURES  
Location/Qualifiers  
1..2823  
/organism="unidentified"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32644"  
/note="Jonesia sp., DSM 14140"  
1..2823  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAD58415.1"  
/db\_xref="GI:26000970"  
/db\_xref="RENTREMBL:CAD58415"  
/translations="MLRTVVAVALTALLATTAGGIAATSTPMPVTAPPSFSSPSGY  
SWSNVEIVGGGFVPGVFNOKDGLAVRTDIGVRLNDSCTGRWILPDIHIGMDWS  
HSGILSLADPVDTNRVYLAAGYSMDWDPDQGAIIIRSDAKGTWKTMLPFRVGNM  
PGRMGERLAVDPNNKVLKLYFAGSNGLWKSVDYDKWGVTSFPAAGYVADASGA  
TYQNGQVAVTDFPTSAKAGQTTIYGVADKQNNVRSSTDGATWQRPQPTGF  
LAQGLVDHKGQQLYATSTGTGPGYDGSKDVWRLDISGQWTRISPISTSSNFAF  
YSLADRNKPDPTIMVVSQVSWMPVYRSTDRKTSPIWELNGSQPRTKQYNHDY  
SGAPWLDGNTAKEPANPKLWMTQSFELDPHNSDRFPFGTGAAGLYGNTLNWDKG  
KVDITVKAQIBETAAQDLAPPGNIGLYSALADIGFTHKDISQPNKYITKNPHH  
DVTYSIDFAESKPTAAVRAGKISIGBTSSVWGVSTDAGETWPKGATPGSVKGGPSIV  
SNASSIIVAPEGAAPRRSTNSGSSWSVSLGPHNAQVADSRVNAANTLXGFVDGKPYH  
STNGGASFTASAFTGPTSGNFRVAVPGRQHLWLAGGVSGSYGWMRSTDGKKNW

KVSVQBGDAVFGKATSSSGYPVIFTSAKIDGVRGIFRSDDGKTKWRINHNQWY  
WNTASITGDDPDEGRVYIGTNGRGLIVGDSSTPPPPGQDSASNGTATPPTPDS  
DHGSGTDTGETSGTSGTSGTAPDSTSPGNTSSCKVRYSTDDGSGTGAATI  
TNTSPSPINGWTLAFTYPSGTQISSVWSATQTLSGRNVLVKNSGNWNPITPAGGSVEFG  
FNGAYTSKNTPPTPSFSLNGVTCALAD"  
1..69  
10..2820  
/product="unnamed"  
ORIGIN  
Alignment Scores:  
Pred. No.: 4,16e-97 Length: 2823  
Score: 2063.00 Matches: 386  
Percent Similarity: 67.21% Conservative: 112  
Best Local Similarity: 52.09% Mismatches: 225  
Query Match: 51.11% Indels: 18  
DB: 6 Gaps: 9  
US-09-917-376-3 (1-740) x AX565635 (1-2823)  
Qy 6 TyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn 25  
Db 130 TATAGCTGTCGACGTTGACATTTGCGGCGGCTTTTGTGCTGGAATTCCTTAC 189  
Qy 26 GluGlyAlaProGlyIleLeuTrpValArgThrAspIleGlyMetTrpArgTrpAsp 45  
Db 190 CAGAAAGACCCAGGCTTGGTGTATGCCCGCACAGATATTGTTGGAGCGTACCGGCTCAAT 249  
Qy 46 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTrp 65  
Db 250 GACTCCACCGGCGCTGGATCCCTTTACGACCATATTTGGTGGGATGACGTGGAGCCAC 309  
Qy 66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaVal 85  
Db 310 TCCGGAATTCCTTGGCCACTGACCTGTGACACCAACCGCGTGTACCTTGACGCA 369  
Qy 86 GlyMetTrpThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGln 105  
Db 370 GGCACGTACTCCATGACGTGGGATCCACAAACGGTGTATTTGCTTCCGCGGACAAA 429  
Qy 106 GlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArg 125  
Db 430 GCGGAGACCTGGGAAAGACGATGTCGCGTTCGTTGGGAGGCAACATGCGCGGTCCG 489  
Qy 126 GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTrpPheGlyAla 145  
Db 490 GGGATGGAGAACCCCTCGCGCTTGACCCCAACAATAATAAGTCTGTACTTTGGTGG 549  
Qy 146 ProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThr 165  
Db 550 GAATCGGTAATGACATATGCAAGTCACTGACTACGGGAAACCTGGGGCAAGTACG 609  
Qy 166 AsnPheProAspValGlyThrTrpIleAlaAsnProThrAspThrThrGly---TyrGln 184  
Db 610 AGCTTCCCAATGCTGTTAATCTACGTCGCC-----GATGCTTCTGGTGGTACACC 660  
Qy 185 SerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerLeuGlyGln 204  
Db 661 GGGGAAACCAAGGTGTGTGGGTAACTTTTGACCCCAATCAGCGGAAACCTGGTACG 720  
Qy 205 AlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSerArg 224  
Db 721 ACAACAAACGATCTACGTAGGTGTAGCAGACAGCAGACCAACGTTTATCATCTAC 780  
Qy 225 AspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIleProHisLys 244  
Db 781 GATGGCGGGCAACATGGCAACGCGTCCCTGCAACACCCACCGGATTCCTGCGGCAAAAG 840  
Qy 245 GlyValPheAspProValAsnHisValLeuTrpIleAlaThrSerAsnThrGlyGlyPro 264  
Db 841 GGAGTGTCTGACCAATAAGGTGACCACTACATCGCTACCTCAGACTACGATCGGCGCCG 900  
Qy 265 TyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrpThrArg 284



NFCXTBPTPEVSPKLGWVGTLEIDPFNSDXMLYGTGAXLYGCDLLTNWDKQNTITI  
 VKAIGLETSVQALISPPVPHLFSALGDIAGRHEDLEKAPNWTTVQNMGTDTDI  
 DPAELNFMVRVGNVDYDKNWPNTRIGFSYDGKSWFQGNTEPQGTSEGTVAAD  
 GSAAVWAPKAGKVCYSDNGKWKVECANVPEALVYSDRVPNKFYAPKNGKFI SAD  
 KGTFTIESPAAGLPIISGNPKTVIEGDIWLGNNGMWHSTDCGYSFVKISGVEDAS  
 IGFGRKABGETYPAIYIYAKINGVRGIFRSDDCDKTWIRINDDKHQPCCANALITGDP  
 RYGRVFNATNGLIGWGEIAYISNIPSATPTPTSPSTPTPTPTPTPTPTPTPTPTPT  
 PTPPT  
 SAGSIIRPFKIVNGSSVDLSRVKIRYWTVDGDKPQSAVCDWAQIGASNVTFNFKY  
 LSGVSGADYILEVGFSSGAGOLQPGKDTGDIQVRFNKNDWSNYNQADWSLQSMTD  
 YGNAKVTLYVDGLVWQEPGGA"

misc\_feature

/note="Region: endoglucanase domain"

/evidence=experimental

misc\_feature

/note="type IIb cellulose binding site"

misc\_feature

/note="pt-linker"

## ORIGIN

## Alignment Scores:

Pred. No.:	5,85e-97	Length:	3262
Score:	2061.50	Matches:	387
Percent Similarity:	65.62%	Conservative:	96
Best Local Similarity:	52.58%	Mismatches:	238
Query Match:	51.08%	Indels:	15
DB:	1	Gaps:	9

US-09-917-376-3 (1-740) x AF078038S1 (1-3262)

Qy 4 GlnProThrThrTrpSerAsnValAlaIle---GlyGlyGlyGlyPheValAspGlyIle 22  
 |||||  
 Db 376 CAGCCATATGTGTGAAGAATGTTAAATTTGAAGTGGAGGAGGTTTATCACTCGGAATA 435  
 |||||  
 Qy 23 valPheAsnGluGlyAlaProGlyIleLeuThrValArgThrAspIleGlyIleMetTyr 42  
 |||||  
 Db 436 GTGTTTAAACCCGGAAGAGAAAAATCTGTTTACGTTAGGACAGATATAGGAGGAGCTTAC 495  
 |||||  
 Qy 43 ArgTrpAspAlaAlaAsnGlyArgTTrpIleProLeuLeuAspTrpValGlyTrpAsnAsn 62  
 |||||  
 Db 496 CGTAGTACAGCGGTGCAATACATGACACAACTTATGAGCTGGGTAGTTTGTATGAT 555  
 |||||  
 Qy 63 TrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysVal--- 81  
 |||||  
 Db 556 TGGAAATTTGGTGGGTGAGAGAGCATTTGACAGATCTCTGTGATCTCTAATCGGTTATA 615  
 |||||  
 Qy 82 TrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArg 101  
 |||||  
 Db 616 TTGGCCCTGCCAGGGGACTTATACCAATAGCTGGACAGATATGAATGGTGTCTATTTTGGCT 675  
 |||||  
 Qy 102 SerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyValAsn 121  
 |||||  
 Db 676 TCTACTGACGAAGAGATACCTTTTGAATCACTCCACTYCCCTTTTAAATTTGGTGGGAAC 735  
 |||||  
 Qy 122 MetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeu 141  
 |||||  
 Db 736 ATCCCRKGAAGGAATCTGGGAGAAAGGTAGCTATTGATCCAAACATAATAGATAATTA 795  
 |||||  
 Qy 142 TyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrp 161  
 |||||  
 Db 796 TACTTAGGAACAAGAGAGMATGGTGTATGGAAGAGTGAAGATTTATGGAGTGTCTATGG 855  
 |||||  
 Qy 162 SerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThr 181  
 |||||  
 Db 856 AAAAAAGTAAAGATTTTCTTAACCCAGGACATATATTGAAGATCCAAATCGCCCAAC 915  
 |||||  
 Qy 182 GlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSer 201  
 |||||  
 Db 916 GATTATTTGAACCATATAACAGGTGTGTCTGGGTGTATTTGATCTCTACAGTGTAGG 975  
 |||||  
 Qy 202 LeuGlyGlnAlaSerIleThrIlePheValGlyValAlaAspProAsnAsnProValPhe 221  
 |||||  
 Db 976 CTGGTGGAGGAACCAAGATAATTTATGTGGTGTGTGACAGATAAAACCAAGCATCTAT 1035  
 |||||

Qy 222 TrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIle 241  
 |||||  
 Db 1036 TATACAAAGACGGTGGTGCAGCATGCGCAAGCATTACCAGGACAGCCAACTGGCTTACT 1095  
 |||||  
 Qy 242 ProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThr 261  
 |||||  
 Db 1096 CCTCAGCGTGCAGAAATTAAGTTCT---CATGGGATGTTATATATATATATATAGTACACT 1152  
 |||||  
 Qy 262 GlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThr 281  
 |||||  
 Db 1153 CAGGGCCCATATATGTTGATATGGAAGATATGAGATATGAGATACATCACTAAGACCGAGAA 1212  
 |||||  
 Qy 282 TrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSer 301  
 |||||  
 Db 1213 TGGAAAAATATAAGTCCCAATGGCTGCTCAAGATACAT---TATTGGATATGGT 1263  
 |||||  
 Qy 302 GlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrp 321  
 |||||  
 Db 1264 GGATTTGGCTGTAGACGCCCCAGAAATCCGAAAGTAGTGGTAGCAGCTCTTAGCTCATGG 1323  
 |||||  
 Qy 322 TrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIleTrp 341  
 |||||  
 Db 1324 TGGCAGATACCTATATTTGGAGAGTATCGATGTGGGAAACGTGGAAATGATTTGG 1383  
 |||||  
 Qy 342 AspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluPro 361  
 |||||  
 Db 1384 GAATGGAATGCTATCTCTAATAGAACATTCATATATACATGATATTTCTGAGCTCCT 1443  
 |||||  
 Qy 362 TrpLeuThrPheGly---ValGlnProAsnProValProSerProLysLeuGlyTrp 380  
 |||||  
 Db 1444 TGGCTAAACTTTGGAAWTACTRATCTTACACACCTCGAAGTAAGCCCAAACTTGGCTGG 1503  
 |||||  
 Qy 381 MetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGly 400  
 |||||  
 Db 1504 ATGGTTGGAACACATTGAAATAGACCAATTAATTCAGATARAATGCTTTATGGAACAGGA 1563  
 |||||  
 Qy 401 AlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIle 420  
 |||||  
 Db 1564 GCTAATTTGATGATGCGATGTTGACAAATGGGACAAAGGCCAAATATACCACT 1623  
 |||||  
 Qy 421 AlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIleSerProSer 440  
 |||||  
 Db 1624 AAAGTCAAGGTATAGGAATTTGAAGAACTTCAGTCCAAGCACTGATAAGCCCACTGTT 1683  
 |||||  
 Qy 441 GlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThr 460  
 |||||  
 Db 1684 GGACCACTTTATTTCAGTGCATCGGAGATATAGCGGATTCAGACATGAAGACTTTGGAG 1743  
 |||||  
 Qy 461 AlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyr 480  
 |||||  
 Db 1744 AAAGCACTAATTTGGACATATGTTTCAGCTAATATGGGACTACTACTGATATTGATTT 1803  
 |||||  
 Qy 481 AlaGluLeuAsnProSerIleValArgAlaGlySerPheAspProSerSerGlnPro 500  
 |||||  
 Db 1804 GCTGAGTTAAATCCAAATTTTATGTTTCTGTTAGGAAATGTTGATAAACAGTGGAAATCCA 1863  
 |||||  
 Qy 501 AsnAspArgHisValAlaPheSerThrAspGlyLysAsnTrpPheGlnGly---Ser 519  
 |||||  
 Db 1864 AATACTAACGAATTTGGTTCTCATATGATGGTGGAAATCATGTTTTCAGGAAATACT 1923  
 |||||  
 Qy 520 GluProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPhe 539  
 |||||  
 Db 1924 GAACCTCAAGGAACACAGCGAAGGTGGTACAGTTGACGCTGCTGAGATGTTAGTGTCTGT 1983  
 |||||  
 Qy 540 ValTrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSer 559  
 |||||  
 Db 1984 GTTTGGGCACCA-----AAAGGACGGAAGATGCTATTTCTACAGATAACCGAAACAAA 2037  
 |||||  
 Qy 560 TrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValIleAsnPro 579  
 |||||  
 Db 2038 TGGGTTGAATCGCAAAATGTAACCTTCAGAAAGCTATTGTATATTTCTGACAGAGTAAATCCA 2097  
 |||||

```

Qy 580 LysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThr 599
Db 2098 AATAAGTTTATTCATTTAAGAAATGCAAAATTTTATATAAGTCTGATAAAGGAAAAACA 2157
Qy 600 Phe---GlnProValAlaLaGlyLeuProSerSerGlyAlaValGlyValMetPheHis 618
Db 2158 TTTATCGAAGTCTGCTGCTGCTGCTTTACCAATATCAGGA-----AATTTCAA 2205
Qy 619 AlaValProGlyLysGluGlyAspLeuTrpLeuAlaLaSerSerGlyLeuTyrHisSer 638
Db 2206 ACTGTTCCAGGTATAGAGGTGATATCTGCTAGTTGGTAATTAATGGATGCGCACTCA 2265
Qy 639 ThrAsnGlyCysSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGly 658
Db 2266 ACTAGGTGGATATTCATTGTGAAATTTTCAGGTCTTCAGGATGCGCAGTATAGGA 2325
Qy 659 PheGlyValSerAlaProGlySerSerTyrProAlaValPheValValGlyThrIleGly 678
Db 2326 TTTGGTAAACCTCTCGAAGGAGAAACATATCTCGCAATTTTATCTTATGCAAAATTAAT 2385
Qy 679 GlyValThrGlyAlaTyrArgSerAspCysGlyThrThrTrpValLeuIleAsnAsp 698
Db 2386 GGAGTAGAGGTATATTTAGATCTGATGATGATGATGATGATGATGATGATGATGATGAT 2445
Qy 699 AspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaLeuArg 718
Db 2446 GACAAACATCAGTTGTTGTTGCGAATGCGACATACATCGAGATCCACGCGTGACGGT 2505
Qy 719 ArgValThrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIle 734
Db 2506 CGAGTATTTGTTCAACAAACGGCTGGGTATATAAATGGGGTGAAAT 2553

RESULT 6
LOCUS AE007608
DEFINITION Clostridium acetobutylicum ATCC824 section 96 of 356 of the
complete genome.
ACCESSION AE007608 AE001437
VERSION AE007608.1 GI:15023819
KEYWORDS
SOURCE Clostridium acetobutylicum
ORGANISM Clostridium acetobutylicum
Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
Clostridium.
REFERENCE 1 (bases 1 to 12732)
AUTHORS Nolling,J., Breton,G., Omelchenko,M.V., Markarova,K.S., Zeng,Q.,
Gibson,R., Lee,H.M., Dubois,J., Qiu,D., Hitti,J., Wolf,Y.I.,
Tatousov,R.L., Sabathe,F., Doucette-Stamm,L., Soucaille,P.,
Daly,M.J., Bennett,G.N., Koonin,E.V. and Smith,D.R.
TITLE Genome sequence and comparative analysis of the solvent-producing
bacterium Clostridium acetobutylicum
JOURNAL J. Bacteriol. 183 (16), 4823-4838 (2001)
MEDLINE 21359325
PUBMED 11466286
REFERENCE 2 (bases 1 to 12732)
AUTHORS Childress,D., Zeng,Q. and Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (24-JUL-2001) GTC Sequencing Center Production,
Finishing, and Bioinformatics teams, Genome Therapeutics Corp., 100
Beaver Street, Waltham, MA 02453-8443, USA

FEATURES
source
1..12732
/organism="Clostridium acetobutylicum"
/mol_type="genomic DNA"
/strain="ATCC 824"
/db_xref="ATCC:824"
/db_xref="taxon:1488"
73..2592
/gene="CAC0919"
73..2592
/gene="CAC0919"
/codon_start=1
/transl_table=11

gene
CDS

```

```

/product="Probably secreted sialidase; several ASP-boxes
and dockerin domain"
/protein_id="AAK78895.1"
/db_xref="GI:15023820"
/transl_table=11
/translation="MNKRIVSMVAGLSIIFTTGVTHISAAANKAAASVQSOGYKWDN
AKIGAGGYPAVIVFKTEKDLIYARTDMGGAYRWDKANNKWIPIITDGFSDMTMLGCE
IATDIPDNRVYIAAGLYTNDQDENAYILSSQDKNTWKYQLPFPKVGNNPFRNNG
ERLQIDPNDKILYGARSNGLMKSEYDGTQWVKVDFPDGTQVQDQPNETADKY
GVWETEDPSTGTGSPQTQMYVGAADTKNNIYVYNDGGKTSVAKGQPKGYLPHHG
ILASDGLYISYNTCGPYDGSQGWKYNKTGWTNITTPAVGDTKSGFGGIVSDA
QNPNNVATLNRMWDEIYRSTDAKTWKPIMDNGVYVNETLGVNLDYSAQPMIDW
KGTGTPPDLVPLKGMWGDLEIDPFNSDRMPYGTATLGTDDLTNWDKGNVDISV
KANGIEACVNDVVPTKGAQLLSAVDDCGFYHDDITKVPKMMTPNPSATTSIDY
AESVPNVVRVGNVYDTSKQDQKCGISYDGGKMFSAKNSISGYKAGTVAAGADAK
TIVMSPREGANAAYSTONGKWTPCGSLPQGAQVRSRVPNPKFYGLFKGKFIYSDA
DTPQSSQGTGLTKGKGIKFTVIGHGDIWAGKDGMLHSTDSAGATFKVSGVDAS
GATVGLSKSDTDDGYPALYMDATIDGTAGIFRSDDEGATWVRINDAHQATGSDPYCITG
DPNKYRFVGTNGRGIVGIDIGSQPTPTSPVTLGVDVNDGIIINGRDVWVLTQYIAG
KQVITDKAADVNDGTVNGRDLMLTQYISGKITSPFSSTK"
complement (2634..2978)
/gene="CAC0920"
complement (2634..2978)
/gene="CAC0920"
/codon_start=1
/transl_table=11
/product="Protein related to MIFH/DOPD protein family,
function in bacteria is unknown"
/protein_id="AAK78896.1"
/db_xref="GI:15023821"
/transl_table=11
/translation="MPFINSTVTVKLDKEKSTLKVGLKIELLIPGKSEDMVMGVFK
DNYLPYFKGEKKDAAFEVIFGSAKSKNKLVSAILSMENELSIKPDCTIYITVD
EVSNNWNGSLF"
3104..4252
/gene="CAC0921"
3104..4252
/gene="CAC0921"
/codon_start=1
/transl_table=11
/product="Metal-dependent amidohydrolase"
/protein_id="AAK78897.1"
/db_xref="GI:15023822"
/transl_table=11
/translation="MITIEKQEEFLKLIGYRRLHKLHPELSMEKYETTKRKRWLYE
NDKVLDPFIEGVVGAIEGHEGHEGHEGHEGHEGHEGHEGHEGHEGHEGHEGHEG
HDFTAILGTVAIILSKRKAELIYGRVRIIFORGEETGKSOVIIEGVLRDVDCIFGM
HNKPDLPVGTGIRSGELMASVDRFEIKVIGSGHAGIPNKICIDPIVVASOIVASMTQ
VWSISPIESVVISITRFNSGTWNVISDSMEGTVTFNNNIEERINLERTAK
GIARALGKIEFWKFSYSPVLENDVRFKDLCEAAEDNGYKVEAKRNILGGSDFSFYQ
TVPSFFVWVGVDGTEWHKANYLNKEDALMAVANYFSDLALKVLSSK"
4594..4863
/gene="CAC0922"
4594..4863
/gene="CAC0922"
/codon_start=1
/transl_table=11
/product="Uncharacterized small conserved protein, homolog
of yfjA/yukE B subtilis"
/protein_id="AAK78898.1"
/db_xref="GI:15023823"
/transl_table=11
/translation="MFDKAALECEDIKKSIEKSTTQLDLSWYGESCTAFKTYLILDR
KLNTYTILOEYSKALYDISETYQNTDGDIAKMSKAVNEASEQ"
5056..6348
/gene="CAC0923"
5056..6348
/gene="CAC0923"
/codon_start=1
/transl_table=11
/product="Predicted membrane protein; CF_1 family"
/protein_id="AAK78899.1"
/db_xref="GI:15023824"
/transl_table=11
/translation="MQSLKVDWNRNGLLKINEAKQGLKSAVDKFSLNKRSRIEILKN
LKRELKVNQLEGEIRKINSINDYMKRFTDVKKCAKSLKTSYGTYYKLLNFGGE
EVERVPVIGIGVDAVESWVRNKFVLTITLTETGAIGVCIITVLEAGITWKA
WDALSAAGGVMSLDNIGTAASAIYQKVERKSEEDSSGVDLFQSFIFGGAGWYSAKILG

```



```

VRNPSTWSSGKTLVDRTSATVTVLNVGHSVTSGLGDKDVEKLPALKABASSANKTLEA
ANNAEKIHDAMWATNKINGINNAKKEVVDLKKLKNVKNFEPYKQFYKDKNLL
KXDNKLPLOQDQLLIQQDQLETFEKNITLAKKNYKNSIRDLNKNYRDIKASLYNIPA
GGTLYNTEDEVQGSFHFNPGEIKPIFEEIKVVLGGER"
/gene="CAC0924"
/6345..6923
/genes="CAC0924"
/codon_start=1
/transl_table=11
/product="Predicted membrane protein"
/protein_id="AAK78900.1"
/db_xref="GI:15023825"
/translation="MIRFOKRMALFELTISILICGDIIEKIIISFGIEFDDILCNLFPI
GLIIGFRIYRWYKROIKKYTELLKEDRLDKIEYIYDIERVGVVYNTAFYIDGVKC
KLRFPQYCKNDNDYEVLEPIYKNGEKFDPAYADKNDYKYSIESTNEITECEBK
AEKKRKRLNEYVNSRMELAKKYBSEAKVM"
7003..8043
/genes="CAC0925"
/7003..8043
/genes="CAC0925"
/codon_start=1
/transl_table=11
/product="TPR-repeat-containing protein"
/protein_id="AAK78901.1"
/db_xref="GI:15023826"
/translation="MENENVRAQELNNRGLIFERNAGDEALKYENKAIKEDENKFE
AVLNKAALCLANNKVDGAVSCYDKLIIVKYNCKAVFGKGNMFFYKDNIKKALELYN
RSIYLGEMSEGVYINLALCMESIGELDDAIKWLNRALIINDQNIITLNKAAVLIKIK
KPYEALEUCNVIKIVDNEEVYHFKAMLIGEMKIEKAPETISRGALLEGQMTLCY
DKAVLPEKQEFKALCTEKALLFDGNNVLLIKKGDLLTYLKKIDEAKSTYDKILE
FHNNMGEVFAKANLQMLLGEHEDABEFFKEIIRLSGBESVLINSYVYRALNKKWA
KTRKQKRLIRKQ"
7998..8357
/genes="CAC0926"
/7998..8357
/genes="CAC0926"
/codon_start=1
/transl_table=11
/product="TPR-repeat-containing protein"
/protein_id="AAK78902.1"
/db_xref="GI:15023827"
/translation="MGKNBEAKYKAEIKKYNFLIIKIYPDYVKLMFFKANCILRID
YKAEELYEYIIDNDALDEVLYLMRARNRISLNKYEAKEEDLNTVILKNPAYRKVIEL
DEOLKEFLSGKSESCYR"
8379..8675
/genes="CAC0927"
/8379..8675
/genes="CAC0927"
/codon_start=1
/transl_table=11
/product="Hypothetical protein, CF-2 family"
/protein_id="AAK78903.1"
/db_xref="GI:15023828"
/translation="MNSDRIIDIEKACDDVVDKLVAGVKVETYNLIHLSNRHIEDG
IKGSADVSIITEYKLVKVMKVENQDNGNDVAANNIVSSFDAEDKWKTKFYR"
8713..9012
/genes="CAC0928"
/8713..9012
/genes="CAC0928"
/codon_start=1
/transl_table=11
/product="Uncharacterized small conserved protein, homolog
of YfjA/yukE B.subtilis"
/protein_id="AAK78904.1"
/db_xref="GI:15023829"
/translation="MVDRCIDIEIDASIFETIKVYSDSKVRLHDIICDLENELRKMD
TWEGDAKKEFDSFTFPGFYSAMKCDMLBELIKELKLVKTSPESLDTEMKNLDKK"
9712..10335
/genes="CAC0929"
/9712..10335
/genes="CAC0929"
/codon_start=1

```

```

/transl_table=11
/product="SAM-dependent methyltransferase"
/protein_id="AAK78905.1"
/db_xref="GI:15023830"
/translation="WGSQVDFDNTAAASNNVIRSEYFEERLKYKVLKSVLSIKDKITADF
CGGTGPFISLALASDANWVGLDNNMLRELKSAHDKLNNIYLKSLTSVTLFDE
SIDAVFINMALHHIKDAEIKEMRYVLKDKGTVVISDVSEHNGWAKEEMFEWLGFG
SNEQIDMWLKNVGFKNIDIENTDLCKGKTSRGEHTETSIFIAKAA"
10353..11498
/genes="CAC0930"
/notes="metB"
10353..11498
/genes="CAC0930"
/codon_start=1
/transl_table=11

```

Alignment Scores:

Pred. No.:	1.34e-93	Length:	12732
Score:	2009.00	Matches:	377
Percent Similarity:	65.41%	Conservative:	107
Best Local Similarity:	50.95%	Mismatches:	240
Query Match:	49.78%	Indels:	16
DB:	1	Gaps:	12

US-09-917-376-3 (1-740) x AB007608 (1-12732)

Qy 3 ThrGlnProTyrThrTrpSerAsnValAlaIleGlyGlyGlyGlyPheValAspGlyIle 22

Db 181 AGCCAGGATATAAGTGGGATATGCAAGATAGTGGTGGTGGATATGTTCCAGCAGTT 240

Qy 23 ValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyr 42

Db 241 ATATTTAATAAACAAGAAAGATTATATATATGCTGCTACAGATATGGGAGGAGCATAT 300

Qy 43 ArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsn 62

Db 301 AGATGGGACACAGGCTAATAAAGTGGATTCCTCAATTTACTGAT-----GGATTTTCTGAT 354

Qy 63 TrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTyr 82

Db 355 TGGCAATGCTTGGTGGCAAGATATACCACTGACCAATAGACACTAATCGTGTATTAT 414

Qy 83 AlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSer 102

Db 415 ATAGCAGCAGGTTGTACACAAATGATTGGCAAGATGAAATGCATATATCTTAGTTTCA 474

Qy 103 SerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMet 122

Db 475 CAGGATAAAGGTAACACATGGGAAAGATACCAACTTCCATTTAAGGTAGGAGGTAATATG 534

Qy 123 ProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyr 142

Db 535 CCGGAAAGAAATATGGGAAAGAAAGCTTCAAATAGATCTCTAATGATGATAAATACTTTAT 594

Qy 143 PheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSer 162

Db 595 CTTGGAGCTGAAGTGGTAATGGACTTTGGAAAGTGAAGACTATGGACAAACATGGTCT 654

Qy 163 GlnMetThrAsnProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGly 182

Db 655 AAGGTTGATAATTTCCCTGATACAGCGGCTATGTACAGAGTCTCTCAAAATGAA-----708

Qy 183 TyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerLeu 202

Db 709 TATACAGCAGATAAGTTGGAGTAGTTTGGGAAACCTTTTGATCTCTCAACAGGACAAAA 768

Qy 203 GlyGlnAlaSerLysThrIlePheValGlyValAlaAsp---ProAsnAsnProValPhe 221

Db 769 GGAAGCCCACTCAACTATGATGTTGGTGCAGCAGATAGACAGGAAACAATATTAT 828

Qy 222 TrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIle 241

Db 829 GTTACTAATGATGAGGAAAGACATGGAGTGCAGTTAAAGGACAGCCAAAGGATATCTT 888



Qy	600	Phe---GlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHis	618
Db	1954	TTTACTCAAAAGTTTCACAAACAGGTCTTCTTACAAAAGGCCAAA---GGAATC---TTTAAAG	2007
Qy	619	AlaValProGlyLySgLUgLyAspLeuTriPLeuAla---AlaSerSerGlyLeuTyTrHis	637
Db	2008	ACTGTTATAGGTCAATGAAGTGATATTGGATTCTGGAGGAAAGACGGATTATGGCATT	2067
Qy	638	SerThrAsnGlyGlySerSerTriPserAlaIleThrGlyValSerSerAlaValAsnVal	657
Db	2068	TCACTGATCTGGTGCACATTCACAAAAGTATCTGGTCTGTATGCTCTCTGATACAGTA	2127
Qy	658	GlyPheGlyLySserAlaProGlySerSerTyProAlaValPheValValGlyThrIle	677
Db	2128	GGTCTTGAAAATCCAAAACAGATGATGGATATCCAGCTATATACATGGATGCTACAAAT	2187
Qy	678	GlyGlyValThrGlyAlaTyArgSerAspAspCysGlyThrThrTrpValLeuIleAsn	697
Db	2188	GATGGAAACAGCGGAATATTCCGTTTCAGATGATGAAGGTCGAATCTGGGTAAGAATAAAT	2247
Qy	698	AspAspGlnHisGlnTyGlyAsnTriPtyGlnAlaIleThrGlyAspHisAlaAsnLeu	717
Db	2248	GATGATGCTCACCAATATGTTCCCCAGACTACTGCATCTAGGATCTCTAACAAATAT	2307
Qy	718	ArgArgValTyrlleGlyThrAsnGlyArgGlyIleValTyrglyAspIleGlyGlyAla	737
Db	2308	GGACGAGTGTGTTGTTGGAACATAATGGAAGAGGTATTGTTTATGGAGACATAGATGGTCT	2367
RESULT 7			
AJ585344			
LOCUS			
DEFINITION	2950 bp DNA linear BCT 11-OCT-2003		
ACCSSION	Clostridium thermocellum xghA gene for beta-1,4-xyloglucan		
VERSION	hydrolyase, strain F7.		
KEYWORDS	AJ585344		
SOURCE	AJ585344.1 GI:37651952		
ORGANISM	beta-1,4-xyloglucan hydrolase; xghA gene.		
REFERENCE	Clostridium thermocellum		
AUTHORS	Clostridium thermocellum		
TITLE	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;		
JOURNAL	Clostridium.		
FEATURES	1		
AUTHORS	Zverlov, V.V.		
TITLE	New cellulosome components of C. thermocellum		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 2950)		
AUTHORS	Zverlov, V.V.		
TITLE	Direct Submission		
JOURNAL	Submitted (09-OCT-2003) Zverlov V.V., Fachgruppe f. Mikrobielle		
FEATURES	Mikrobiologie, TU Muenchen, Am Hochanger 4, Freising, 85350,		
source	GERMANY		
gene	Location/Qualifiers		
RBS	1..2950		
CDS	/organism="Clostridium thermocellum"		
	/mol_type="genomic DNA"		
	/strain="F7"		
	/db_xref="taxon:1515"		
	346..2887		
	/gene="xghA"		
	346..350		
	/gene="xghA"		
	359..2887		
	/gene="xghA"		
	/note="belongs to GH74 and contains a cellulosomal		
	dockerin module"		
	/codon_start=1		
	/evidence=experimental		
	/transl_table=11		
	/product="beta-1,4-xyloglucan hydrolase"		
	/protein_id="CAE51306.1"		
	/db_xref="GI:37651953"		
	/translation="MWKFKTSKTKAAVFAAFAAATAIFGPAISQAQTSVPYKWDNVV		
	IGGGGFMPIGVNETEKDILYARADIGGAYRWDPSSTETWIPLLDHFQMDQESYGVGE		
	SIATDPVPRNRVIVAGMYTNDLWLPNNGAILRSTDRGETWERTILPFKMGNGMPGRSM		



```

Db 2219 GTGGGATTAACTTTACCGATACAAAGGACCCGAGCTTCCCAAGTCGGTTAAAC----- 2272
QY 615 ValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAala---SerSer 633
Db 2273 ---AAGATAAAGCCGTACCGGCAAGGAAGACATGATGGCTTCTGCAAGAGAGGC 2329
QY 634 GlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSer 653
Db 2330 GGATTGTGGAGGTCCACTGACGGTGGATATACGTTTGAGAAACTCTCCAAATGTTGACACA 2389
QY 654 AlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheVal 673
Db 2390 GCTCATGTGGTACGCTTCGGAAGGACGACCCGGACAGGATTACATGGCGATTACATT 2449
QY 674 ValGlyThrIleGlyValThrGlyAlaTyrArgSerAspAspCysGlyThrThrTrp 693
Db 2450 ACCGTTAAATTCACATGTTTAGGATCTTCGTTCCGATGATGCCGCAAGACATGG 2509
QY 694 ValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAsp 713
Db 2510 GTGGGTATCAACGACGACGACGATATGCGCTGTTGATCTGCAATAACAGGTGAC 2569
QY 714 HisAlaAsnLeuArgValTrpIleGlyThrAsnGlyArgGlyIleValTyrGlyAsp 733
Db 2570 CCAGAGATATACGGAGCTGTATATATTGCCCAACGGAAGGATATTGTTTACGGCGAA 2629

RESULT 8
ABI292929
LOCUS ABI292929 2579 bp mRNA linear PLN 08-AUG-2000
DEFINITION Agaricus bisporus mRNA for CEL6 protein.
ACCESSION AJ292929
VERSION AJ292929.1 GI:9755896
KEYWORDS cel6 gene.
SOURCE Agaricus bisporus
ORGANISM Agaricus bisporus
Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
Agaricales; Agaricaceae; Agaricus.

REFERENCE
1
AUTHORS Morales-Almora, P. and Thurston, C.F.
TITLE Molecular analysis of the cellulolytic genes in Agaricus bisporus
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 2579)
AUTHORS Morales-Almora, P.
TITLE Direct Submission
SUBMITTED (07-AUG-2000) Morales-Almora P., Microbiology, King's
College London, 150 Stamford Street, London, SE1 8WA, UNITED
KINGDOM
FEATURES
source
1..2579
/organism="Agaricus bisporus"
/mol_type="mRNA"
/strain="D649"
/db_xref="taxon:5341"
/clone="PMRT152"
/dev_stage="mycelium"
/country="United Kingdom"
1..2579
/gene="cel6"
28..2448
/gene="cel6"
/function="putative cellulase"
/codon_start=1
/product="CEL6 protein"
/protein_id="CAC02964.1"
/db_xref="GI:9755897"
/db_xref="GOA:Q9P4T8"
/db_xref="SPRENEL:Q9P4T8"
/translators="MPSRVKFWLTLILGTFSLVYGOAASQSYRNQNVKIGGGGFPVPG
IVFNPSERGLAYARTDIGAYKLNADDTWPLDFFADDSRWNYWGVDAIATDVQPNR
LVLATGLYNSWDPNSNGHILISSDYKTFDAAPLPFKLGGNMPGRMGKIVLVDPNLN
SILYFCARSGNGLWKSNGRSWSKVTFTDGTGSPVDPDSDTGLNSDKIGIAWVTID
KASGSGSATPRTFVGVANKGNSIYISTNGSSSWAVAGQPTSLPHKGLVLSERA
LITYSDGGGPGDYTGSAVKYKSIDITGAWTDITPVSGSDLFFPGGLALDITQNGTVM

```

```

VAALNSWPDQIIFRSTNGGASWTPLMDWAVIPTLTKYYSYNAALAPWIGPNIVDTP
GNLQIGWMSLSIDPFDNSHLYGTGTETIYGSRLDKWDSAHNTYIISLADGVBETS
QVALISPSGPPVLSAIGDVGCFVHQSTRAPOFTNPKWSTADIDFAGNPNPVV
RIGDTSGSKOVAISSDYVTWQHFAPDNVCGKVAISADADIILPWTNGNGVMV
SRNOATFNI FYGASGKTFYVSTDKTKTFSAGSLGSAVSVDITVHPVSUGDIIWASTD
KGLFHSSTDGATFSAISGITOAWGALCAPRSTGYPAPVPAANAANTGGEVAYFRSDRG
VNWVKINDAAHFGAASANCMAADPRVYRVIYTGNGRGI FYDVGAGSAPPPTSSATS
TTSYVTTTTRSTTTTTRSTTTTSSGNGSPVSGGCGGWTGTTCTVGTGFOCVAQNP
FYSOCIP"
28..96
/gene="cel6"
97..2445
/gene="cel6"
/product="CEL6"

sig_peptide
mat_peptide

ORIGIN
Alignment Scores:
Pred. NO.: 2,85e-81 Length: 2579
Score: 1753.50 Matches: 341
Percent Similarity: 60.97% Conservative: 112
Best Local Similarity: 45.90% Mismatches: 249
Query Match: 43.45% Indels: 41
DB: 8 Gaps: 10

US-09-917-376-3 (1-740) x ABI292929 (1-2579)
QY 1 AlaThrThrGlnProTyrThrTrpSerAsnValAlaIle---GlyGlyGlyGlyPheVal 19
Db 94 GCAGCTTCTCAGTCTTATAGGTGGCAGATGTCAAGATCGCGGTGGTGGTGGTTC 153
QY 20 AspGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGly 39
Db 154 CCGGCAATGCTTCAATCCTTCCGAAAGGACTGGCGTATGCTCGCACTGACATCGGA 213
QY 40 GlyMetTyrArgTrpAspAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGly 59
Db 214 GGAGCTACAGCTGAAT---GCTGATGATACTTGGACGCTCTACTTATTCGCTGAT 270
QY 60 TrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsn 79
Db 271 GATTCTCGTGGGAATATTGGGGCGCTCGATGCTTTCGAACCTGATCCCGTACCAACCAAC 330
QY 80 LysValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIle 99
Db 331 CGTCTTTACCTAGCAGCTGGGTGTATACCACTCTGTTGGACCCAGTAGAACGCCATATT 390
QY 100 LeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGly 119
Db 391 CTCATATCAAGTGATTTATGTTAAACAATTCGATGCCGCGCTTTTGCATTCGAAGCTCGGT 450
QY 120 GlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsn 139
Db 451 GGTAATATGCTGGCGTGGAAATGGGAGAGCGCTCTCGTCTCGATCCCAATTCGACAGC 510
QY 140 IleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAla 159
Db 511 ATCCTTTATTTTGGCGCTCGCAGTGTAAAGCTCTCTGGAAGTCAACAATTCGAGTCGG 570
QY 160 ThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAsp 179
Db 571 AGCTGGAGCAAGGTAAACAAGCTTCACTGACACTGGGTCTTTCTGTTCCCGCATCCATCAGAT 630
QY 180 ThrThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSer 199
Db 631 AGCACCGGCTTGAATTTCTGATAAGATTGGGATTCGATCGATGGTGTACCATCGATAAGCGTCA 690
QY 200 SerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspPro---AsnAsn 218
Db 691 GGCTCTCTGCTCGCTACCTCCCGTATATTTTGTGGCGTGTTCGGAACAGGCGCTCAAAC 750
QY 219 ProValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThr 238
Db 751 AGCATTTATATATCTACGAATGGCGGTCTTCTTGGAGTGGCGTTCGAGGTCAACCAACT 810

```





```

Qy 615 valMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGly 634
Db 1865 GCCATC---CGCGCGCATCCCTCCATCCCGCGGCGATCTCGGGCGTCGACCGCAAGCGG 1921

Qy 635 LeuTyHisSerThrAnGlyGlySerSerTrpSerAlaIle---ThrGlyValSerSer 653
Db 1922 CTCGGCATCCAGGATACGGTAGTAGCTTCACGAGATCGCAGCGCGGTCACCGCG 1981

Qy 654 AlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTrpTrpAlaValPheVal 673
Db 1982 GCGTGGAGCTTCGGTTTGGCAAGCGCTCCACGGCGAGCTACGTCGTACTACGGG 2041

Qy 674 valGlyThrIleGlyGlyValThrGlyAlaTyArgSerAspCysGlyThrTrp 693
Db 2042 TTCTTACCATTCCAGCGTCCGCGGGCTGTTCAAGAGCAGGAGCGCAGCACCACCTGG 2101

Qy 694 ValLeuIleAsnAspAspGlnHisGlnTyArgGlyAsnTrpGlyGlnAla-----IleThr 711
Db 2102 CAGGTGATCTCGGATCGTCGACGGATTCGGCTCG---GGTCCGCCAACGTCGTGAAC 2158

Qy 712 GlyAspHisAlaAsnLeuArgArgVal-TyrIleGlyThrAsnGlyArgGlyIleValTy 731
Db 2159 GCGCACTGCAGACTACGCGCGGCTGTTCTGCGGCGACGAGCGCGCGCATCTCTA 2218

Qy 731 rGlyAspIleGlyGlyAlaProSerGly 740
Db 2219 CGGCAATCCCGAGCGGAGCTCGCGGCG 2246

RESULT 10
AY040839
LOCUS
DEFINITION
Aspergillus niger endoglucanase C (eglC) gene, complete cds.
ACCESSION
AY040839
VERSION
AY040839.1 GI:19879405
KEYWORDS
Aspergillus niger
ORGANISM
Aspergillus niger
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
REFERENCE
1 (bases 1 to 3959)
AUTHORS
Hasper, A.A., Dekkers, E., van Mil, M., van de Vondervoort, P.J. and de
Graaff, L.H.
TITLE
EglC, a new endoglucanase from Aspergillus niger with major
activity towards xyloglucan
JOURNAL
Appl. Environ. Microbiol. 68 (4), 1556-1560 (2002)
MEDLINE
21914059
PUBMED
11918668
REFERENCE
2 (bases 1 to 3959)
AUTHORS
Hasper, A.A., Dekkers, E. and de Graaff, L.H.
TITLE
Direct Submission
JOURNAL
Micro-organisms, Wageningen University, Dreijenlaan 2, Wageningen
6703HA, The Netherlands
FEATURES
Location/Qualifiers
1..3959
/organism="Aspergillus niger"
/mol_type="genomic DNA"
/db_xref="taxon:5061"
<1038..>3841
/gene="eglC"
join(<1038..1137,1185..1329,1373..1571,1618..1858,
1908..1961,2007..>3841)
/gene="eglC"
/product="endoglucanase C"
join(1038..1137,1185..1329,1373..1571,1618..1858,
1908..1961,2007..3841)
/gene="eglC"
/codon_start=1
/product="endoglucanase C"
/protein_id="AAK7227.1"
/db_xref="GI:19879406"
/translation="MRRTPAALVAGYLLDSVHAASQAYTKNVVTGGGGFTPGIVF
NPSAKGVAVARTDGGAYRLNSDDTWTPLMDWANNNSNHDWGDIAIATDPVDTRVYV

```

## ORIGIN

```

Alignment Scores:
Pred. No.: 3,27e-74 Length: 3959
Score: 1620.00 Matches: 352
Percent Similarity: 55.57% Conservative: 107
Best Local Similarity: 42.62% Mismatches: 255
Query Match: 40.14% Indels: 113
DB: 8 Gaps: 20

US-09-917-376-3 (1-740) x AY040839 (1-3959)

Qy 1 AlaThrThrGlnProTyrThrTrpSerAsnValAlaIle----- 13
Db 1095 GCCCATCGCAGGCATATACCTGGAGATGTGGTCAC-AGGGGGTATCACAATCTTGC 1153

Qy 14 -----GlyGlyGlyPheValAspGlyIleVal 23
Db 1154 AATATGAGGGATTGCAAGCTAATGTATGACGGCGCGAGGCTTCACGCCGGGTATCGTC 1213

Qy 24 PheAsnGluGlyAlaProGlyIleLeuTyValArgThrAspIleGlyGlyMetTyArg 43
Db 1214 TTCAACCTTCAGCTAAAGGTGCGCATATGCTCGCACCGACATCGCGGGTGCATATCGG 1273

Qy 44 TrpAspAlaAlaAsnGlyArgTyrIleProLeuLeuAspTrpValGlyTrpAsnAsnTrp 63
Db 1274 TTGAAC---TCGGACGATACATGACCCCGTTGATGGACTGGCCCAACACTCTAACTGG 1330

Qy 64 -----GlyTyr-Asn-----GlyValVa 69
Db 1331 TACGTAAGCTTCTCTTACGAGATGCTAGGCTCACTGTCCAGGCACGATTGGGGCATTCGA 1390

Qy 69 lSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTyThr 89
Db 1391 TCGATCGCGACCGACCCCGTCGATACCGATCGGGTATATGTTGCAGTTGGCATGTACAC 1450

Qy 89 rAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTr 109
Db 1451 CAACGACTGGATCCCAATGACGGCTCCATCTCGCGGTGACCGACCGAGCGGCGATACCTG 1510

Qy 109 pGlnIleThrProLeuProPheLysLeuGlyAsnMetProGlyArgGlyMetGlyGl 129
Db 1511 GGAAGACAGCAAGCTTCCTTTCAAAGTTGGAGGAATATGCCCGGTGGGGTGTGGCGA 1570

Qy 129 u-----ArgLeuAlaVala 134
Db 1571 GGTAAAGTCTATATCTGGGTTTGTGTAGCATACACTGACCGCTGCAGCGCTCTGGCTGTGG 1630

Qy 134 spProAsnAsnAspAsnIleLeuTyPheGlyAlaProSerGlyLysGlyLeuTrpArgS 154
Db 1631 ATCCGATGACCAACAGCATCTTGACTTTGTGCGCGAGTGGCAATGGCTTGTGGAAA 1690

Qy 154 erThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyTr 174
Db 1691 GCACGGACTACGGCGAGACATGGTCCAATGTGACTGCCTTCAAGTGGACCGCACCTATT 1750

Qy 174 leAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrpVala 194
Db 1751 TCCAAAGATTCCAGCTCGAC-----TACAGCTCTGATCCCGTGGGAATTCCTGGGTCA 1804

```





[illegible]





/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/clone="002-167-F04"

## ORIGIN

## Alignment Scores:

Pred. No.: 2 42e-71 Length: 2724  
Score: 1560.50 Matches: 330  
Percent Similarity: 60.24% Conservative: 126  
Best Local Similarity: 43.59% Mismatches: 252  
Query Match: 38.66% Indels: 49  
DB: 8 Gaps: 22

US-09-917-376-3 (1-740) x AK10506 (1-2724)

Qy 6 TyrThrTrpSerAsnValAlaIle---GlyGlyGlyGlyPheValAspGlyIleValPhe 24  
Db |||||  
127 TACTCATGGAAGATGCCAGTTTGGAGCGCGGTGGCTTTGTGCCGGCATGTATTC 186  
Qy 25 AsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyMetTyrArgTrp 44  
Db |||||  
187 CATCTACCCAGGCTGGCGTCTATGCTCGAAGCATATAGGAGGCTGTACAGGTGG 246  
Qy 45 AspAlaIleAsnGlyArgTrpIleProLeuLeuAspTrpVal-----GlyTrp 60  
Db |||||  
247 AAC---AGTGATGACTCGTGAGCGCCCATCACCGACTCCATCCCGAATGCCAATGG 303  
Qy 61 AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 80  
Db |||||  
304 CACAAATGG-----GAGTGGATGCGGTGCTCTGATCCCGAGGCCCAATGTG 354  
Qy 81 ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 100  
Db |||||  
355 GTGTACTTCGCGCGGCTGTGTACACCAACTCTCTGGGATCTTAGCCCGCGCCATCTAC 414  
Qy 101 ArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGly 120  
Db |||||  
415 AAGAGCAGCAGCAGGCTGGACCTGACACCTTTCACAGCTCAGCTTCAAGGTGGTGGT 474  
Qy 121 AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle 140  
Db |||||  
475 AACATGCTGCGCGCGCATGGGAGCGCTGTGCTGTGATCCCAAGAACTCCAACTT 534  
Qy 141 LeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThr 160  
Db |||||  
535 GTGTATTTCGCGCACCTCTGGAACGGCTCTGGAAGAGCAGCAGCGCGGCAAGACC 594  
Qy 161 TrpSerGlnMetThrAsnProAspValGlyThrTyrIleAlaAsnProThrAspThr 180  
Db |||||  
595 TTCAAGAAAGTTCGACTCTTCAAGGCTGTGGTACCTTACATCCCGGTCCCGCAGCAGC 654  
Qy 181 ThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSer 200  
Db |||||  
655 ACCGGCTACAAACAGCAGCATCAACGGCTGACCTTTGTGACCTTTGATCGACCTCCGCA 714  
Qy 201 SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAsp---ProAsnAsnPro 219  
Db |||||  
715 ACCATCAATGTTGGTCCATGCGCGCATATTGTTAGGACACCGCCGATACGTACCGCTCG 774  
Qy 220 ValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGly 239  
Db |||||  
775 CTTTACGTTTCTGAGGATCGCGCAGCAGCAGTGAACGCGGTCAAGGCGCAGCTCGCAAA 834  
Qy 240 PheIleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSer 259  
Db |||||  
835 TACTTCTCTAATAGTCAGGCTGCACGCCACCAAGCGCGCATGTACATCTCTTACGCC 894  
Qy 260 AsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSer 279  
Db |||||  
895 GACGAAACGGGCCCTTACGATGGTCTCTGGCGCTGTCTGGCGATACGGCTCGCGACC 954  
Qy 280 GlyThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGly 299  
Db |||||

Db 955 GGCAGTGGAGCGACATCACTCTCTCCGGTAGGACCTGACC-----TACGGA 1005  
Qy 300 TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIle 319  
Db |||||  
1006 TTCGGCGGCTCTCGTGACATGAAGAAGCCAGGAACCTCTGGTGTGGCAGTCTCAAC 1065  
Qy 320 SerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArg 339  
Db |||||  
1066 TCTTGGTGGCTGATCGCAGATCTTCGCGAGCACCGACTCTGGCAAGACGGTGTCAAG 1125  
Qy 340 IleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAla 359  
Db |||||  
1126 CTCTGGCAGTGGGTAACTATCCGAGATGATCTACTATTACGGCATCTCGACTCCCAAG 1185  
Qy 360 GluProTrpLeuThrPheGlyValGlnProAsn---ProProValProSerProLysLeu 378  
Db |||||  
1186 GCCCCTGG-----ATCCGCAAGAACTTCTCTCGATACCGACTTCCAAGTGGCTG 1233  
Qy 379 GlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGly 398  
Db |||||  
1234 GGCTGGATGATCGAGTCTCTCGAGATCAACCGACTGACAGCGACCATCTCTGTACGGC 1293  
Qy 399 ThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIle 418  
Db |||||  
1294 ACCGACTGACCTCTACGGTGGCCAGCCTGACCAACTGGGACCAACAGCGCAACATC 1353  
Qy 419 HisIleAlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIleSerPro 438  
Db |||||  
1354 ACCATCCAGTCTCTGGCGCAGCGCATCGAGAGACCGCTGTTCAGAGGTGGCCTCGGC 1413  
Qy 439 ProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHis----- 456  
Db |||||  
1414 CCCGCGCAGCAGCTCATGCTGCCACGCGCGACATCTCCGGCTTCACTTCGCATCG 1473  
Qy 457 ---AlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrGly 475  
Db |||||  
1474 GCGCAGCAGCTTGGAACTTCCAAAGCGGTCTGGGACCAACCCCAAGTTTCAACCTCA 1533  
Qy 476 ThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAsp 495  
Db |||||  
1534 GCGAGCTGATACGCGCGCAACTCGCGCGGAGGTGTGGCGTGGGCAAC---GAT 1590  
Qy 496 ProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrp 515  
Db |||||  
1591 GCGCGGCTCAG-----ATGTCGCGACCTCGACCGATGGCGGCTCATCATCG 1638  
Qy 516 PheGlnGlySerGluProGlyGlyValThr-----ThrGlyGlyThrValAlaAlaSer 533  
Db |||||  
1639 -----AACGTGTACCCCGCGCGCACACGAACAGCGCGGTGCGGTGCTTACTCT 1692  
Qy 534 AlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAla 553  
Db |||||  
1693 GCCATGTTGACATGCTGTGTGTCACCGCGCAGCGGTGGC-----GTGTACCG 1743  
Qy 554 ValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArg 573  
Db |||||  
1744 TCCGTGAACAGGCGCAGTTTCTCCGCGCTCTCTGTCAGAGCGAGCTCTCGTGGCC 1803  
Qy 574 SerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSer 593  
Db |||||  
1804 TCGACAAAGCGCAACAAACAGTACTTCTACCGCGCATCCGCGAGCAGCTTCTACGTCTCC 1863  
Qy 594 ThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaVal 613  
Db |||||  
1864 GTCAGCGCGCGTGACCTT-----GCGCGCGCGGTCTCGCTCAGCGCGCGCACG 1914  
Qy 614 GlyValMet---PheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSer 632  
Db |||||  
1915 TCGATCCGCGACATCGCGCCCGCCAGCAGCGCTGGCAAGCTTTATGTGTGCGCAAC 1974  
Qy 633 SerGlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThr---GlyVal 651  
Db |||||  
1975 GCAGGCGTCTTCCACGAGCAGGACTTTGGCGAGACCTTTTACCAAGACTTTCGCCGCGCTG 2034



```

/note="SAV2464"
/codon_start=1
/transl_table=11
/product="putative ATP-dependent helicase"
/protein_id="BAC70175.1"
/db_xref="GI:29606112"
/translation="MKPSPLELHAAVAVGTRPQVMTAEVAEALDDSSHLV
QAGTGGKSLGLVLPALAGHGVIVATIALQRLVRLPTVDAHPHLLRRPPEF
AMLKGSNTLCRLHREGLFQFEAAAPTSLKQDQLRLRWSDETETGD
RDLTPGSDRAMAQVSVSSRECLGATKAYGAECFAEMSRERAKLAEVVTNHALLA
IDAEAPVLPOHEVILVDEAHVLSRVGTGELTPOGVNRAVRAAKLVNEKAD
QLTAAEGPERLMLALPGHLEIPEDLGYALMALDAAETVISAIGATKSDQVED
AVRQALASVESVHDVAERTINGSEVDVWYERHDFRGLASRVASGLLREKLPFA
DRSVLTSATLKLGGDFNGGASLGAPEGTQGDLLPQWKGVDGSPFLPKQGLIV
AKHLSRPARDGRDMLDELITQAGRTGLFSSMRQAQAAEELSRIPFEFPL
LOQETLGLIKNFAADPETCLFTLSLQGVDPVPGPSVOLVWMDKIPPRPDPLMS
ARQKAVEDAGGNGFMAVAAHAALLMAQAGRLVRASGRGVAVLDQRLATARYGSY
LKASLPDFWTTDRNQVRSLSLTALIDENAKTETVTEVAEVTBEAAKSEGA"
complement(7376..8509)
/note="SAV2465"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC70176.1"
/db_xref="GI:29606113"
/translation="MPPTDASTDAEPATSDAPARIAVSARSAADSEDTLELRPDELV
ALPFAEEAAGGEEAAGDAGVGEETAGGEEDAVGAEPEAGRKDAGPEAQGTATA
APSVPARAGATAGSDGSDAVDPAAPFAAATVPSASRVPSAPSVTAPAT
APAPAPPSGDTKALAGDSVSTRAADLLGRVEDNGPTVTPAGTQFLAPVPERDL
PLLGRWNDPAVAAFVKLAGDESVTQHLRAQGGGDSRVPCLGLEGTPMSTWEIYR
ADDSLRHYPARPHDTGTHLLIGGVADRGRLGSLTLLRAVADLVLDRRPSCARVAE
POLRNTSSVSFAFAGFRFAEVDLPDKRAALMVRDRPLRDLI"
complement(8549..10609)
/note="SAV2466"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC70177.1"
/db_xref="GI:29606114"
/translation="MNAPTSDGSRTHERRAGHDTOLPTVOGQVAVPASAQAAPPE
TQAPLAPATQVVFVEVTRPVQASSEVPPVGRARESERLRTGHTDPLDHPAQTHA
QAAAVENLURKVRNGLVPPDDGTILRIPLLATGALLVPHVHSTDPHRRFGLPYLA
DAFQAPADAVTVALLAREAAAGRLRPRPAHSDVHDCAVDSADLVGRVADSVRT
AVEIDRRARPADPEOLLAAEQALLHPLHPTPKSREGLSEAEARLYSPELRGSPA
LHMLAVAPVLVATNSAWTERGRVPVPAEOLTAIAGDGLPMPDGPALPLHWPOLIRETR
HREAAALLDAGLIRDLGPHGPAAWHTSSVTRTGCAPAMKLSLGLRITNSRENIL
RKLHGRVEVHRLRGLGQWQARHPGFDIVRDPALVADDDQGNAPVGLDVMIRHN
PFAPTDVSCVAGLVSRRPMPQTEQPPVMSRLAEIIRLAGRTGRPRGAVATEWF
LRYLEQVPRVLMDSBAGIALEAHQONTLLLDPEGWPTGGRYRDNQGYFFTSRRA
ELEEARLPGIHEHSDTFVSDVTERFAYVILGINNVGLIGAFGSHLAERQLLAAFR
RFLADVATGPARTPLPTRLDPSVPLRCKANLLTRLHGLDELVLGVPDVTQSVVTIAN
PLHS"
complement(10775..12229)
/note="SAV2467"
/codon_start=1
/transl_table=11
/product="putative class III aminotransferase"
/protein_id="BAC70178.1"
/db_xref="GI:29606115"
/translation="MAVTSKSAVSLKREAEAGHAHEGILRRQSARESAARTYARALP
IVPRARGITIEAGDRYRLDCLSGATLALGHNHFWLEAIRKVLDSGAPLHVLIDIA
TPVKDAFTLPTLPLAGLADRARVQFCGPGAGTDAVEAPKLVRTATGRSGLLAFTGA
YHGMTAGALASGATDGRVARLPYQDVCECPGIGEGEAGELGASWTESLLDDVKSQ
VPLPAGMILEPVOGEGGVIIPADAWLRMRATIDRSIPLIADVETQVGTGRFWAV
EHSVVFDVWMLSKAIGGSLPVLAVVVRDLDVWQGAHAGTGRNQLMAAAGETASP
VYREHLAERAOQLTGLARMLSLATEFPFCVGEVRGRGLMIGVELVRAEDAEVASP
GSDGFASLDTAGAGPRPAAPELAAVORECLRRGLIVELGGRHASVVRLLPPLTISDE
QATAVLDRDLADVAEAVARGAERSRGRFSERAG"
12761..13993
/note="SAV2468"
/codon_start=1
/transl_table=11
/product="putative secreted protein"

```

```

/protein_id="BAC70179.1"
/db_xref="GI:29606116"
/translation="MRPIRPLFAARRGRSARRRTPSVLGAVALATPALATATACDSGD
DTKASSESPSASSAGDDKIKIPDDIRDKLKERGIDIDKWKNGAWKNWNRDLDLEAN
Alignment Scores:
Pred. No.: 4,61e-61 Length: 299550
Score: 1405.00 Matches: 308
Percent Similarity: 54.57% Conservative: 104
Best Local Similarity: 40.79% Mismatches: 257
Query Match: 34.81% Indels: 86
DB: 1 Gaps: 21
US-09-917-376-3 (1-740) x AP005031 (1-299550)
Qy 6 TyrThrTrpSerAsnValAlaIleGlyGlyPheValAspGlyIleValPheAsn 25
Db 161381 TACCGCTGGCGCAACGCGCTCATCGGGGCGACCGCTTCGTACCGCGGTGCTCTTCCAC 161322
Qy 26 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrpAsp 45
Db 161321 CCCTCCGTACGCGGTCTCGCTACGCCCGGACCGACATCGGCGGCGCTTACCGCTGGGAC 161262
Qy 46 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr 65
Db 161261 GACCGCGCGCGCGCTGGACCCCGCTCATGACCACTCGGCTGGGACGACTGGAACTTC 161202
Qy 66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 85
Db 161201 CTCGGCTGCGAGGCGATGCGCGTGCACCCCGACCGACCGCGGCGGCTCTACCTCGCGGTG 161142
Qy 86 GlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGln 105
Db 161141 GGCACCTACGCCCGCTCGTGGCGGGCAAC---GGCGCGGTCTCGCTCCGAGGACCGC 161085
Qy 106 GlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArg 125
Db 161084 GGCGCCCACTGGACCCCGCACCGCTGACCGTGAAGCTGGCGGCAACGAGGACGCGCGC 161025
Qy 126 GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAla 145
Db 161024 GGCSCCGGTGAGCGACTCTCTCGTCCACCCGCGGACGACGACACCTCTGGCTGGGACG 160965
Qy 146 ProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThr 165
Db 160964 ---CGGCACGACGGGCTGCTCAAGTCGACCGCCGCGGCGCCACTTGGGACGCGCGACC 160908
Qy 166 AsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSer 185
Db 160907 GCCTTCCCG-----GCCAAGCGCGAACTCC 160884
Qy 186 AspileGlnGlyValTrpValAlaPheAspLysSerSerSerSerLeuGlyGlnAla 205
Db 160883 TCCGGCGAGGAGTCTGTGTTTC-----CTCGTCCGCGCC 160851
Qy 206 SerLysThrIlePheValGlyValAlaAspProAsnAsnPro-----Val 220
Db 160850 GGGCGCACCGTCTACGCGCGGTGGGGTGCACGCGGACGCGACCTCGGCGACGCGGAACCTG 160791
Qy 221 PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe 240
Db 160790 TACCGACGCGCGGAC---GGCACGACCTGGGGGCGCGTCCCGCGCGCGCTCCGCGACC 160734
Qy 241 -----IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAla 257
Db 160733 TCCGCCAAGGTCCCGCTCCCGCGCGCGTACGACACGACACCGCGAGCTGTACGTGACG 160674
Qy 258 ThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerVal 277
Db 160673 TACGGCGACGACCCGCGCGCGCGCGCGCGTCCGACGCGGAGCTGCACAGCTGCGTACC 160614
Qy 278 ThrSerGlyThrTrpThrArgIleSerProVal-----ProSerThrAspThr 293
Db 160614 ::::::::::::::::::::::::::::

```

Db 160613 GCCACGGGACGTGGACCGAGGTACCCCGGTGAGCGCGGGGAGCAGCAGCGACGGC 160554  
Qy 294 AlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIle 313  
Db 160553 TCGCGGACACCTTGGCTACGGGGGTGGCGTGCAGCGCCCGCGGACCCCTC 160494  
Qy 314 MetValAlaThrGlnIleSerTyrTrpProAspThrIlePheAspSerThrAspGly 333  
Db 160493 GTCGTCTCCACCAACACCGCTGGCGGCGGACGCGGACACGCGTCTTCGCGTCCACGACGGC 160434  
Qy 334 GlyAlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyr 353  
Db 160433 GCGCGTACCTGGAGCTCCCTCAAGGAC-----GCCGCC 160401  
Qy 354 ValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProVal 373  
Db 160400 GTGTTGACGTGTCGAGACTCCCTTCTCGACTGGGGCGAGC----- 160356  
Qy 374 ProSerProLysLeuGlyTyrMetAspAlaMetAlaIleAspProPheAsnSerAsp 393  
Db 160355 ---AAGCGAGTTCGCTGGTGGATCCAGCGCTCGCGGTGCGACCGTACGACTCCAG 160299  
Qy 394 ArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAsp 413  
Db 160298 CACGTCGTGACGGGACCGCGGACCTCTACGGCACCCCGGACCTCAAGCGCTGG--- 160242  
Qy 414 SerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluThrAlaValAsn 433  
Db 160241 -----GCACCGCGATCCGCGGCTGGAGGAGCGCGTGGCGC 160203  
Qy 434 AspLeuIleSerProSerGly---AlaProIleLeuSerAlaLeuGlyAspLeuGly 452  
Db 160202 CAACGTATCTCGCCCGGTGGGAGGACACCTGATCAGCGGACTCGGGGACATCGT 160143  
Qy 453 GlyPheThrHisAlaAspValThrAlaValProSerThr---IlePheThrSerProVal 471  
Db 160142 GTGATGTACACGAGCGGTACCGCGCTCTCGTGGCGGCGCATGGGACGACGACCCCGTG 160083  
Qy 472 PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAla 491  
Db 160082 TTCGGTTCGGCGGACGAGTCTCGCGAGCGCGCGGCGGCGGTATGTCGTCGCGACG 160023  
Qy 492 GlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly 511  
Db 160022 GCTCGGCGGACACCGGCAACGGC-----GGTACTCCACGACGCGC 159981  
Qy 512 GlyLysAnTrp-----PheGlnGlySerGluProGlyValThrThrGlyGlyThr 529  
Db 159980 GGGCGGACCTGGGCGCCTTCGAGCGCCCGCCGACATCGCCAGGACGACCGGACCG 159921  
Qy 530 ValAlaIleSerAlaAspGlySerArgPheValTrpAla-----ProGlyAsp 545  
Db 159920 ATCGCCACGAGTCGCGGCGGACACGTGCTGTGTCTTCGTCGACATGGGACGCGC--- 159864  
Qy 546 ProGlyGlnProValValThrAla-----ValGlyPheGlyAsnSerTrpAla 561  
Db 159863 -----ACGAGTACCGCGGCGCCACCGCTCGACGGACACCGCGGAGTGGTCC 159816  
Qy 562 AlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThr 581  
Db 159815 GAGGTCTCTCTTCGAGGCGGCGCACACCGGTGCGCGACCGCGGATCCGACGCGC 159756  
Qy 582 PheTyrAla-----LeuSerAsnGlyThrPheTyrArgSerThrAspGlyValThr 599  
Db 159755 TTCTACGCTACGACTTCGACATGAAGCACTATACGCGGACGACTGACGATGGCGGTTCG 159696  
Qy 600 PheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAla 619  
Db 159695 TTCACGCGCGGTGGCGGAGTCCGCGGCGGACGACCGGATTCAGCTGGTCCGCG 159639  
Qy 620 ValProGlyLysGluGlyAspLeuTyrAlaAlaSer---SerGlyLeuTyrHisSer 638  
Db 159638 GCGCGGAGCAGACGCGGACCTGTGGCTCTCCGCCAAATGGAACGGGCTCTACGCTCC 159579

Qy 639 ThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGly 658  
Db 159578 ACCGACGGCGGGACACCTCGCCAGGATCGACAGCTGTGGCGCTCGTACACCCCTCGC 159519  
Qy 659 PheGlyLysSerAlaProGlySerSerTyrProAlaValPheValValGlyThrIleGly 678  
Db 159518 TTCGCAAGGGCGGCGGCGGCGGACCTACCGCGGATCTACAGTCTGGCTCGCGGAG 159459  
Qy 679 GlyValThrGlyAlaTyrArgSerAspAspCysGlyThrThrValLeuIleAsnAsp 698  
Db 159458 ACCATCACCGCGTCTACGCTCCGACGACGCGCGGACGATGGTCCGGATCAACGAC 159399  
Qy 699 AspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArg 718  
Db 159398 GACGCCACACGAGTGGGGTGGATCGCGAGCGCGTCTGGTGTACCGCGGATCCACGGC 159339  
Qy 719 ArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAsp 733  
Db 159338 CGGGTCTACCTCGCCACCAACGCGGCGGCGGATCCAGTACGGGGAG 159294

## RESULT 14

AE011809 14520 bp DNA linear BCT 29-MAY-2002  
LOCUS Xanthomonas axonopodis pv. citri str. 306, section 187 of 469 of  
the complete genome.  
ACCESSION AE011809  
VERSION AE011809.1  
KEYWORDS GI:21107967  
SOURCE Xanthomonas axonopodis pv. citri str. 306  
ORGANISM Xanthomonas axonopodis pv. citri str. 306  
Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
Xanthomonadaceae; Xanthomonas.  
REFERENCE 1 (bases 1 to 14520)  
AUTHORS da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R.,  
Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida  
Jr., N.P., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C.,  
Camargo, L.E.A., Camarotte, G., Cannavan, F., Cardozo, J.,  
Chamberg, P., Ciapina, L.P., Cicarelli, R.M.B., Coutinho, L.L.,  
Cursino-Santos, J.R., El-Dorry, H., Faria, J.B., Ferreira, A.J.S.,  
Ferreira, R.C.C., Ferro, M.I.T., Formighieri, E.F., Franco, M.C.,  
Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite  
Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A.,  
Madeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Meidanis, J.,  
Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M.,  
Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira Jr., H.A.,  
Rossa, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinola, L.A.F.,  
Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezza, R.I.D., Trindade dos  
Santos, M., Truffi, D., Teai, S.M., White, P.F., Setubal, J.C. and  
Kitajima, J.P.  
TITLE Comparison of the genomes of two Xanthomonas pathogens with  
differing host specificities  
JOURNAL Nature 417 (6887), 459-463 (2002)  
MEDLINE 2202145  
PUBMED 12024217  
REFERENCE 2 (bases 1 to 14520)  
AUTHORS da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R.,  
Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida  
Jr., N.P., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C.,  
Camargo, L.E.A., Camarotte, G., Cannavan, F., Cardozo, J.,  
Chamberg, P., Ciapina, L.P., Cicarelli, R.M.B., Coutinho, L.L.,  
Cursino-Santos, J.R., El-Dorry, H., Faria, J.B., Ferreira, A.J.S.,  
Ferreira, R.C.C., Ferro, M.I.T., Formighieri, E.F., Franco, M.C.,  
Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite  
Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A.,  
Madeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Meidanis, J.,  
Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M.,  
Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira Jr., H.A.,  
Rossa, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinola, L.A.F.,  
Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezza, R.I.D., Trindade dos  
Santos, M., Truffi, D., Teai, S.M., White, P.F., Setubal, J.C. and  
Kitajima, J.P.  
TITLE Direct Submission

JOURNAL Submitted (28-NOV-2001) Departamento de Bioquímica, Universidade de Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900, Brazil

FEATURES

source Location/Qualifiers  
1. .14520  
/organism="Xanthomonas axonopodis pv. citri str. 306"  
/mol\_type="genomic DNA"  
/strain="306"  
/db\_xref="taxon:190486"  
/notes="pathovar: citri"

gene

87. .2342  
/gene="cela"  
/notes="synonym: XAC1770"  
87. .2342  
/gene="cela"  
/notes="identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer/Genemark"

CDS

/codon\_start=1  
/transl\_table=11  
/product="cellulase"  
/protein\_id="AAM36634.1"  
/db\_xref="GI:21107968"  
/translation="MTAVIDPPTKRRBSAAHMLTAVCLTLVLLFWTSEADERTGP  
DPQWMSVAGCGGFTGVLFFHFAERDLYARTDVGAYRWDAQAQWALTDLWGLD  
DNLWGLDAPVDPNPNALYLAAGTVMHERRAGNAVLRSDRGKQFERADLPFLKGG  
NQLGRANGERLAYDPHDGRLVLLSGDRGLWSDRRGAHWARVEGFPADALAGATARN  
HVGSEQAVGTAFFVFDAAAGHAGATPRIYGVSTQATSLYSEDAGRTWSAVAGOPK  
GLRPMHMGHOGWLSYDGRPGDPLMAGGLWNTDTCGRWRETSPIPOPATGDF  
GKAVAPDHPHPOVLLAPRPRETPRDEPRSGDGRSVPLLAAAPGHAAPWTAH  
ATPHWIGALIDPFDNSHATFVYGYGIWASRNLTQTSQQPLOWWQDRGLETVPLD  
LISPMAGHILSALGIDGRHDAIDTAQLYGLPRLTNGESIDGAGAPQWVVRSGT  
VDRNRNEALYSODGTHAAFASEPFRGQAGATIAAADSQVWVVDQGGVWRT  
GDFGRQWRQVGLPDTAVVADVADQWVLAADRSGLRYSSESDGASFDTCQOQS  
PARDEARQPLDPPWRAGVYLAAPTLGVWRQNGQLRLTSKSDARSIGIKALRA  
GAPPALYLAGRVAGDGIFRSDDEGAHRRINDAHRFGKPYSTGDPRIAGRVYAT  
GGRGIFGDRP"  
2431. .4347  
/gene="XAC1771"  
2431. .4347  
/gene="XAC1771"  
/notes="identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer/Genemark"

gene

9201. .11573  
/gene="XAC1774"  
9201. .11573  
/gene="XAC1774"  
/notes="identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer/Genemark"

CDS

/codon\_start=1  
/transl\_table=11  
/product="sialic acid-specific 9-O-acetyltransferase"  
/protein\_id="AAM36635.1"  
/db\_xref="GI:21107969"  
/translation="MTASFVRRGCLGLVLVSPCAWAVPTPLLLADGAVLQRDQMP  
VWGSSPNAIAVSDGKRAVADATGQWVRLPAHAGGPPYLRVQGGDELQVRD  
VLVDVWLAGGQNMWELQAQSDGPAVAANDAOQLRQPKPKWSVQPAQLTGE  
WKAATPANAGEFTAVGVFFFAKELURASTGVPIGVNSTWGGSAIEAMWDAASIGLNADO  
NQGAIEALKORDAAQAATGKRLARWPKVEDENPQWREAFDSDWDSIPVTKWBS  
GYDMDGIAWIRYITTLAARAKAGITLGVQDIDSDTITTYNQCGQVSTGKQWNLPRV  
YQVPAALAKGVANHIAVRVEDSGGGMGHPDEQFVQTSAGAKRALDGKWFPAARV  
VSLTDKNOLPVLVYNMHPLOFPFVKVGIWYQETNATDTGAVRYEQFAADIRGMP  
RABRGDKTLPFLVWAKAGDKGELSPWALLRESQSQTLLALPQFGQAVIIDIGNP  
TDHPNKRVDGHRILAAHVAHYGETLVSAVFEKASPDGKAVLGEDLOQSALQV  
RGGAVQGRIRIAGDQFHPATQIDGDRIVRSDAVAAPVARYGSENPPDANLIN  
RDALPVSPPRTFW"  
4344. .6101  
/gene="XAC1772"  
4344. .6101  
/gene="XAC1772"  
/notes="identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer/Genemark"

gene

11570. .13063  
/gene="XAC1775"  
11570. .13063  
/gene="XAC1775"  
/notes="identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer/Genemark"

CDS

/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAM36636.1"  
/db\_xref="GI:21107970"  
/translation="MTRLSMSHSPVARVIASSPARRRRLPLLSALALAAACQBSGE  
QDDKAPAAVATAGAAQTPTAAAPAAQTTPMPQFVSKDGKHALWVGAPFLILGAQ  
VNNSNYPGIMDKVPMQVLPNTVQVPIAWEQVEPEGKDFSVDTLLNQAREHK

VLRLVLFATWKNNGPOYADPHWVXTNQRPRPVVTRECKSIGSLSPHAQATLADARKA  
FVAPMQLKKVDPQHTVIMIQPENEAGTIGSVRDYFSPNAQKLDFQGVDELIGKWKQ  
PGTWAQHLKADAEIFHAWSIGRIDYDIAQAKAEYPLPMYVNAALRGPFPNGPQGY  
ASGPTNDVLVDVWKAAGPHDILLAPDIYMPYRMVTVLERYAPDNLFAVETGNRP  
EYARLYPTLGHNGIGSAFGIDTRYSNYPLGAKHVNNEETLAPFALGFQAVSMGWRP  
FAKASGKLGHTAEERPCQAOELKNDMSATITYGVQFWEKGEPCGNPEPIGAAL  
IAELSPDEFLLTGHYARTVTLHPASDATANMIYDRVESEGVEGNDWKFORNWNQDQTDY  
GNPSSPFLQTKIKLATYK"  
5999. .9109  
/gene="xyls"  
/note="synonym: XAC1773"  
5999. .9109  
/gene="xyls"  
/notes="identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer/Genemark"

/codon\_start=1  
/transl\_table=11  
/product="alpha-xyloridase"  
/protein\_id="AAM36637.1"  
/db\_xref="GI:21107971"  
/translation="MEIPAQLQEMRPDRLRQFFORAAAQQAQAGNLQISAPATTPRRR  
RDGIASPPCAAPAPRHSKGVIMOTIRISAGAMPANRRTSLAQCLALSALLANA  
AHQEVKADQGVTVPSAKGAAVRLQVVDAGIIRVSADPDGDFARSPSLMRVPQG  
DTAFQVAEQSDVQTKGVKTSISTVDGHSVPADANGKPVLSVAGRSFAPLNVG  
KQYLSVRQFQSPDEALYFGQHQQGMNKGKRNELQNNIDNAVLYLSSRNYGL  
LWNNSTIRLGDPRGLQPLPKTLTLYDAKAGALTAARYAINGKHILERRSEVNYQY  
LSDITPKKAITKDKSRMQVTWEGIEALTGGEHTFSLYSSEYAKLYVDGKLVYDR  
WRONNPNWHEFLDLBPGRHTVKVWDLIDPSYALLHRDPLPAAKADLSLVAEEY  
QOMDYFVSADSDQAVAGYRALTGKSTMLPKWAYGFQWQRERVKSDODELVAGAEY  
RKRLSLDNVLVDSWYMPENAGSHDFDPQHPDDGMVKAHVDMHAQIMISWPKFY  
PTTAYKELDAAGFMKRNVEGELMIGKYGKNSFYDYPSEKAQAIYWRQINEKLS  
KGFDAWMDADEPDVHNLDAERKARTTNALGSSTFYNSYPLPHTHVGTVGRRA  
DDKVFILSRKAGTGNVAVWMSGDIVSRWDMRDQISGVNNAISGLPNMTDFIG  
GFAVEKEDQPAHLPEWRELNRWFQFAGVPIFRSHGQFPYREIWNIAPEGVPFY  
ESMAYNRLSALPYLSLAGDTYQRDQVIMRGMMDFPNDKYLDINDOYLFGPAF  
LVAPVTRGATSRQVLYPAGSWLEFATGKRYEGQCSIEAAPPIERMELFVAGSIVP  
TGPVQYEVYQVADAPLVVVTGADQGFSLYEDDGKGYEKGEFESRPLVWVWQAGE  
LSIGKREGSWTGMQAKRTINRVFDGPRDDAGALAPKTDASTIQYDGKFPVSLQRIAS  
GKARRR"  
9201. .11573  
/gene="XAC1774"  
9201. .11573  
/gene="XAC1774"  
/notes="identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer/Genemark"

/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAM36638.1"  
/db\_xref="GI:21107972"  
/translation="MLTRRELCKATGAIVLGVAPAAQAQATQASKTTALPPVAAAE  
LQWYREPAWNEVVALPVNGRLGAMVGGIAHERLOLNEDTLVAGGPYDSTSPDALA  
ALPOVRALIFAGRYAEAEKADAKLSRPLKMPYQPLGDLILLDFRADGIDSIRYRL  
DLDTAVATITFRSGGAVHRRVFCQAQCIIVRLSCDRPGGISURVGDISTQGEVT  
AEPGLDFRNGSGFAGIEGRFALRVLPQVSGKLSQVRDLRLIDRAADEVYLLLSA  
ATSQVRDADVDGDPALTAARLAKAALDIPFALLRAHLADHQLRFRVADILGSSAV  
QLPTDERVORFAGNDPALAALHYOGRYLLICSSRPGTQANTQCIWNLDMPWPWES  
KYTININWNYWSEANALHECVLEPMLFDLAQTCTHTARAIVDAPGVWVNTD  
LWRQAGIDGACSLWPMGVLVQLQDLWRWDYGRDAYLSKVYELFKGAAEFFVATL  
MRDPQTGAMVNTFMSPENHPFGAAVCAQSPMDQLLDLFAQCIASMLKGLIDQAL  
AQDUALREQLPNRIKAGQLOQWQDMQAPETINHRVSHVYALHPSSQINLRDT  
PELAAARSLRLEIRGDNATGIGWRNLNWARLADGEHAYRILQILIPRTYPNLFD  
AHPFQIDGNFGGTGIGITMELQSWGSGVFLLPALPKAWPRGSRGLRVGGSVDLE  
WEGSRLOARLHSDRGGRYQLSYAGATLDLELGAGRTQOVGLNNRLVTQ"  
11570. .13063  
/gene="XAC1775"  
11570. .13063  
/gene="XAC1775"  
/notes="identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer/Genemark"

/codon\_start=1  
/transl\_table=11  
/product="D-xylulokinase"





```

Qy 647 -----AlaIleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAla 663
Db 2067 CAGTTCGCGCGGATCAAAATCAGATGAAGCGCGTCGCGTGGCATCGGCAAGCGGCTA 2126
Qy 664 ProGlySerSerTyProAlaValPheValValGlyThrIleGlyGlyValThrGlyAla 683
Db 2127 CGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGATC 2186
Qy 684 TyrArgSerAspAspCysGlyThrThrTyrValIleAsnAspAspGlnHisGlnTyr 703
Db 2187 TTCGCTCCGACGACGAGCGCGCCATTCGCGCGCATCAACGACGATGCCATCGCTTC 2246
Qy 704 GlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgArgValTyrIleGly 723
Db 2247 GGCAAA---CGGTATACGTCGACCGCGGACCGCGGATTCGCGCGCGGTACTTTGCC 2303
Qy 724 ThrAsnGlyArgGlyIleValTyrGlyAsp 733
Db 2304 ACCGGCGCGCGTGGCATCTTCTATGGCGAT 2333

RESULT 15
AE001712
LOCUS AE001712 13278 bp DNA linear BCT 02-JUN-1999
DEFINITION Thermotoga maritima section 24 of 136 of the complete genome.
ACCESSION AE001712 AE000512
VERSION AE001712.1 GI:4980799
KEYWORDS
SOURCE Thermotoga maritima
ORGANISM Thermotoga maritima
REFERENCE 1 (bases 1 to 13278)
AUTHORS Nelson,K.E., Clayton,R.A., Gill,S.R., Gwinn,M.L., Dodson,R.J.,
Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A.,
McDonald,L., Uterback,T.R., Malek,J.A., Linher,K.D., Garrett,M.M.,
Stewart,A.M., Cotton,M.D., Pratt,M.S., Phillips,C.A.,
Richardson,D., Heidelberg,J., Sutton,G.G., Fleischmann,R.D.,
Eisen,J.A., Fraser,C.M. et al.
EVIDENCE for lateral gene transfer between Archaea and bacteria
from genome sequence of Thermotoga maritima
JOURNAL Nature 399 (6734), 323-329 (1999)
MEDLINE 99287316
PUBMED 10360571
REFERENCE 2 (bases 1 to 13278)
AUTHORS Nelson,K.E., Clayton,R.A., Gill,S.R., Gwinn,M.L., Dodson,R.J.,
Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A.,
McDonald,L., Uterback,T.R., Malek,J.A., Linher,K.D., Garrett,M.M.,
Stewart,A.M., Cotton,M.D., Pratt,M.S., Phillips,C.A.,
Richardson,D., Heidelberg,J., Sutton,G.G., Fleischmann,R.D.,
White,O., Salzberg,S.L., Smith,H.O., Venter,J.C. and Fraser,C.M.
Direct Submission
TITLE Submitted (01-JUN-1999) The Institute for Genomic Research, 9712
JOURNAL Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source
1. 13278
/organism="Thermotoga maritima"
/mol_type="genomic DNA"
/db_xref="taxon:2336"
150..1985
/gene="TM0300"
150..1985
/gene="TM0300"
/notes="similar to GB:AE000657 percent identity: 48.20;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="oligopeptide ABC transporter, periplasmic
oligopeptide-binding protein, putative"
/protein_id="AAD35388.1"
/db_xref="GI:4980800"
/translations="MKRFLFVAVLFTFAAFQFLNVPREQTIVVPGPKTANPGDFN
VIAGWKDPHRIQQLLEPLMDPVRGEIINALAETGIYNEDFTNIVKLRGVYV
SDGVPTADDVFTVEMCKYPEMARHMEFNEIDKYKEDDYTVVPELKKPNTRPHT

```

```

YFIDRWGWMPKPKVFEKVEDPVGFTFNPVPVGTGPPYVLKDYDPAGYVWLWEREDWQ
RTPTGFMPLGMPKPKYVLFYAYETDEKITLAFQNLQDAAGIGPQALKALLKRSKYVKL
YSDPFFPLVDVPVGGITINTAKYFPNNKRYRWALADIVKLSSETEGNNLISAIH
IPPLTYRKYLYEPLLEWLNQVNGVKVYDVPDLKIAEVAKR/GWNPVDDPEKIKL
IGVSGWKKHLPDVAEQLLSVGKDKNGKWL/PNGTKIQ/LGNGNDLVAFGVAQO
WREGIDVDVITTYISGNEGFVLTGNWPAPEPGWGHADLHRTLVEFHSKYQNKP
LGEQGHVSRWFDERLDKIIDEMEKLDWNPRIEFGREGLKILAELEPTIPIGTIGV
LTALVNDYDYNPTLENPFYWHWSQFYKLLTFLBPTGRK"
2050..3054
/gene="TM0301"
2050..3054
/gene="TM0301"
/notes="similar to GB:Pyro h percent identity: 59.88;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="oligopeptide ABC transporter, permease protein"
/protein_id="AAD35389.1"
/db_xref="GI:4980801"
/translations="MKFIKSYLIPRLIQYVVMVIGITVLFILPRLSPVDPIEQILNT
LRIONTPTEAVIKMREVMELVGLKGTWEQVNLKRLKLGDFGSPMSVYYPASVIS
LRNSLPWTIGLTLTTLISVIGLVIGALIGYKESFGSKLTGTGLMAINIPYVIF
ALTLLLECYVFPPIFPGGSGVQKISLSWSFILDVVKHGTLPALSLILGFAAGNL
GTLKLVNREEDIVVYARMWGLRNSKILFSTVLRNAMDQVITGLSGLGGFGGWMM
TEILTYPGIGTLAYNAVANDYNVMIASIFSLIAVATALLILDLLYPLIDPRIKYR
"
3073..3942
/gene="TM0302"
3073..3942
/gene="TM0302"
/notes="similar to GB:Pyro h percent identity: 57.09;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="oligopeptide ABC transporter, permease protein"
/protein_id="AAD35390.1"
/db_xref="GI:4980802"
/translations="MKRILRIFDWMEDLRFKYAVTVLTALVLLSILSIFSYPDPYK
VYLTNPQPSLEHILGTGQDIFWLLTFLSKNSILATIAAFSRIITIGMLSG
YLGTDRILMALTDTPMWPVEPLLLLSLKNVLSILPILGILGIVFGWAGDARVI
RSMILSRERFVTSKSGMRTFVEVGFDPVPLIPVISGFGISGMAIGAEIIVLA
ILGTRVEIPTLGSMSFMHMINFQAMLLGYWVWLTPITVAVFLFTALYTLISUSINEYL
DPRTRMQRIGAVK"
3945..4931
/gene="TM0303"
3945..4931
/gene="TM0303"
/notes="similar to GB:U08186 PID:473270 GB:AE000512 percent
identity: 76.00; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="oligopeptide ABC transporter, ATP-binding
protein"
/protein_id="AAD35391.1"
/db_xref="GI:4980803"
/translations="WEKTAVENLRVYFIKSGDKTTSIRAVDNLVHLHFEFNEVYGIA
GESCGKSTLLKVLFGELPKPLVDGEVWFKLDEEYRINSMSLDQLKKRWKYSY
VPQGSVNLNPNRKIRLFDVIBEHTSMSREAFSRIESFEALGPLVRVLDSDPYQ
LSGMQRQRTALATVNLNPKVIFADETPTALDVVQVGVQVQLLKHKEQKNTLVIT
HDMGWHARLATRIGYVAGRLIEAEETHEIFKNPLHPTFRFLNSLPFGDRSRKEGT
PGSPPSLANLPSCGPFHPRCPADQICMEMSPDLIEVARNHKVACWKVGVER"
4928..5878
/gene="TM0304"
4928..5878
/gene="TM0304"
/notes="similar to GB:Pyro h percent identity: 65.51;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="oligopeptide ABC transporter, ATP-binding
protein"
/protein_id="AAD35392.1"

```



```
/db_xref="GI:4980804"
/translation="MRVSVRNLTKI FRVGSIFSLKKLVADNVSVFISGSEIFSLAGE
SCGCTTAKILGLEPTNGEVLYDGIKLVDFENEMHFLKKVQTVQNPFFSTFNP
RKVEDTFLKRVKMANTKDEALRIEKLKLVVGLNFKRDEIKGRYSEFSGQLORIS
IARALLTNPFLVADPVSMDASLRISIVNLFRLDKRFGISILVITHDLSYAVVA
DKLAINFSGSIVEMGAQDILDPFPHTYKLLKSSIPVDPPEYENDENINLSSTEREE
FRANGCKFASRCEPVSICKRESVPPDPVDFNVRVVKWLYK"
gene
5894..8017
/genes="TM0305"
5894..8017
/genes="TM0305"
/notes="similar to GP:3560020 percent identity: 55.11;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="endoglucanase, putative"
/protein_id="AA035393.1"
/db_xref="GI:4980805"
/translation="MLRSFLILFLAILGVVFGATFEWKSVEINGGFPVPGIIFHPASP
GLLYARDVGGLYRWDEETKRWKQDFLRRDQSDYMGVLSVALDPSDPKRIYAMTKG
YTQWAGYGAIISEYGETWTIVNLDKYGIKVGNGEDGRNAGERLQVDPFSSVLPM
GTTKGLWSKSDGKWKWDSFSTSVTFVLFDEKSGEKSSTPPIRFVCSSEPKGIF
VEDGTTWVNLPLNDLIPLRGKHGDIYVTLNAGLPNGATRGAVMKYVIADOK
WIDVTPMKDGFYGGIDVQENVVIVSTDRWYPHDEIFISLNGGTRWPKLEKAFENI
NKAPITKIDLPNHWISDKIDDPDMNRAIFTTGYGVWVYELKKSFEFGKPKVWFENI
RGLEETVVLQVPPIIGERPLLSAIDWNGPRLSDTPSSMYKPLKWTSLGIAFAVO
NSKFVAVHTYVPELISYEDGGINWREIETVPEGITDGRSLSLAYNDGKTLVWSPA
NHEVIVSSDKGSKWKAISVPEPFYPPASDPNPSKFIYIPKNGDDELISDKGKS
FMKAGLPSPDNWVSLYSPVLAOPREGDILWALQNGLYRSKOGGITPERLGNVDI
AVYIGAPKPGDTPAIVLNGVNGVYGFIMSTDEGKTMWRINNDKHQFGWYHMG
DNNEFGRIPLGTGRGIIIVEKKEE"
8014..9363
/genes="TM0306"
8014..9363
/genes="TM0306"
/notes="similar to GB:M29877 SP:P04066 PID:1335066
PID:178409 PID:182779 percent identity: 56.91; identified
by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="alpha-L-fucosidase, putative"
/protein_id="AA035394.1"
/db_xref="GI:4980806"
/translation="MTSMKPRYKPDWESLREHTVPKWDFKAKFGIIFHMGIVSVPGWA
TPTGELCKYPMDAWFPQNPYAEYNSLRIKESPTWEYHVKTYPGENPEKPADLFTA
EKWDPEWADLLFKGAKAYVPTTKHHDGFLMGTKYTFDPSNFKPRDLVDLAKA
VREAGLRFYVSGGDLWFTTPIRYPEDLSYIRPNTYEDYAYKQVMDLVLYLP
DVLMDNWPGEKEDLKYLFAYYKHPGSGVNDRWGVPHWDFKTAELYHVNYPGDLP
GYKWEPTRGIGLSFGVNRNRPHEMLSVLEOLVTVLVVSKGNLNLNVGPKDGTIP
DIQKELLGLGELWLYKGDALYCTSWERCCAKTEGTEIRFTKCNRIEPIELGPT
GKXIVEDNLNAGTVRHFUTGERLSFKNVGKNLEITVPKKLETSDISITLVEAVEE"
9360..10781
/genes="TM0307"
9360..10781
/genes="TM0307"
/notes="similar to GB:I42023 SP:P44779 PID:1005445
PID:1220891 PID:1204865 percent identity: 56.91;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="L-fucose isomerase, putative"
/protein_id="AA035395.1"
/db_xref="GI:4980807"
/translation="MKGRYRVGLISFSDGREYVHRELEGIVKVFPEIATKLEBGEV
EVIKAEIVKMSIAKREARLAEAGEVTFINYVMAFPHTVLASKFAPGFLIFS
NINPOYPMVAMLSAAGAEODCTKHVRMMGNIKDEKLVKWMAFIRAGSTYKFLRGO
RYGVGGRSMGYVTPVNDLVNKKIEGVQVHIDQEIYVRSOLIPDERARKGLMTE
ERKAVHYDKXYLTPKLELQIKSYHARVIREMELDFVGIKQGLETFEHTVMDVT
ETFLNDPYDWEGBEPIVCATEADSDGALTMQIFKLIATKPIVLFADVRHVSEYDVID
LCNSGNHATYFAARSFDDPENMKKVFYPTQTFYPAGGAAGVHIAAPKGVTLGRLETV
DGRYFTVVPVGFVDFEGEKNYEIAQSIQNNWPHAFLEKMETPIDEFELAKYSNNHIGV
YGDYVEIKTFCEIASIDFLMK"
complement (10798..13092)
```

```
/gene="TM0308"
complement (10798..13092)
/gene="TM0308"
/notes="similar to PID:1881669 percent identity: 63.77;

Alignment Scores:
Pred. No.: 4.83e-50 Length: 13278
Score: 1160.50 Matches: 264
Percent Similarity: 51.53% Conservatives: 275
Best Local Similarity: 35.15% Mismatches: 89
Query Match: 28.75% Indels: 89
DB: 1 Gaps: 26

US-09-917-376-3 (1-740) x AE001712 (1-13278)
Qy 6 TyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn 25
Db TTTGAGTGGAAATCGGTGGAATCAACGAGGAGGATTGTACCAGGAATAATATTCAT 6013
Qy 26 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrpAsp 45
Db CCTGCTAGTCTCTGGCTTTTATATCGGAGACGGATGTTGGTGGACTTTACAGTGGGAT 6073
Qy 46 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr 65
Db GAGGAAACAAAGAGCGTGAAGCAATATTATTGATTTTGTGAGAAGAGATCAGTCCGATTAC 6133
Qy 66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnIleValTrpAlaAlaVal 85
Db ATGGGAGTGTGGAGCGTCTCGATCTCTAGTATCGAAGAGAAATCATACGCGATGACA 6193
Qy 86 GlyMetTyrThrAsnSerTrpAspProAsnAspGlyValAlaIleLeuArgSerSerAspGln 105
Db GGAATATACATCTAGGATTGG---CCAGATATGAGGCATCTTATCTCGGAGATTAC 6250
Qy 106 GlyAlaThrTrpGlnIleThrProLeuPro-----PheLysLeuGlyGlyAsnMet 122
Db GGTGAAACGTGCAACATTTGTAACCTTCGCAAGTAGTGGGATAAAGTTGGTGGGACGAA 6310
Qy 123 ProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyr 142
Db GATGAAAGAAACGCTGGTGAAGATTGAGGATAGATCCAAATTTTCAGTACGCGTCTTTT 6370
Qy 143 PheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSer 162
Db ATGGGTACA---ACGAAATATGCTGTGGAATCAGAAGATTTTGGAAAAAATTGGAAA 6427
Qy 163 GlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGly 182
Db AAGTGGATAGTTTTCCTCAACC----- 6451
Qy 183 TyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerLeu 202
Db -----AGTGTAACTGTTGTGCTATTTCATGACAGAAAGAGGAGAAAAA 6493
Qy 203 GlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrp 222
Db GGAAGTCTCTACTCTAGAAATTTTGTGTGATGTTTTCAGAACCCCAAGGGG---ATCTTTGTA 6550
Qy 223 SerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyValAlaProThrGlyPheIlePro 242
Db ACTGAGATGTTGGACCACTCTGGATGTTCTTCCAAACCTGCCGACCATCTCATTCCTCA 6610
Qy 243 HisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGly 262
Db TTGCGTGGAAAAATACAT-----GACGGGATCCTTTACGTTACACTTTTCCAATGCTTTG 6664
Qy 263 GlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrp 282
Db ----- 6665
Qy 283 ThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGly 302
Db -----
```

6725	Db	TTACGATGTTAACCAATGAAGAGAT- - - - -TTCCGATATTGTGGA	6766
303	Qy	LeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrp	322
6767	Db	ATAGATGTTCAAGAG- - - - -AACGTGGTGATAGTATCCACACTCGATAGGTGGTAT	6817
323	Qy	ProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIleTrpAsp	342
6818	Db	CCACACGATGAGATATTATTTCCCTCAACCGAGGAGAAACGTGGAGACCTCTCTTCGGAA	6877
343	Qy	TrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrp	362
6878	Db	AAGCGCAAAATTTT- - - - -GATATCAACAAGCGCCCTGG	6910
363	Qy	LeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGlyTrpMetAsp	382
6911	Db	ATAAAA- - - - -CATTTAAATCCT- - - - -CACTGGATATCT	6940
383	Qy	GluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThr	402
6941	Db	GAC- - -GTCAAGATAGATCCATTTCATCATGAATAGGCGTATTTTCCAAACAGGTTATGGT	6997
403	Qy	LeuTyrAlaThrAsnAspLeuThrIlystProAspSerGlyGlyGlnIleHisIleAlaPro	422
6998	Db	GTGTGGGTGACTTATGAGTTAAAAAAGTCATTCGAAGT- - - - -ATFGGAAAA	7045
423	Qy	MetValLys- - - - -GlyLeuGluGluThrAlaValAsnAspLeuIle	436
7046	Db	CCAGTGAATGGNATATTCGNAATTCGAGGTTAGAGAAACAGTTGTTTGGCACTGGTT	7105
437	Qy	SerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHis	456
7106	Db	CCCCCATAGGAAAGGCGCTCTCTCTCTGCAATCCCGAGTGGGAGGATTCCGTCAC	7165
457	Qy	AlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThr	476
7166	Db	GAAAGTTTGGACACCCCATCATCTATGTAC- - -AAACCCCTAAAATGGACGCTCTCT	7222
477	Qy	SerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspPro	496
7223	Db	GGTATAGCTTTTCCCTATCGAAATTCMAAATTCGTGGCAAGTGTGATACATAC- - -	7276
497	Qy	SerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPhe	516
7277	Db	ACCTATCCT- - - - -TTCTTGCTTACTCCGAGGATGGCGAATCAACTCGAGA	7324
517	Qy	Gln- - -GlySerGluProGlyGlyValThrThrGlyGly- - -ThrValAlaAlaSer	533
7325	Db	GAGATAGAAACTGTCCCGAGGGAATAACTAGCGCGGAGGTTAAGTTTGGCAGTCTCG	7384
534	Qy	AlaAspGlySerArgPheValTyrAlaProGlyAspProGlyGlnProValValTyrAla	553
7385	Db	AACGATGGAAAAACACTGGTCTCGAGTCCAGCGAAT- - -CATGAGGTAAATTGTTCA	7438
554	Qy	ValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAla- - - - -AsnAla	570
7439	Db	AGCGATAAAGGAAAAAGCTGGAATAAGCCATTAGTGTTCAGTTCAGAAATTCAACTAT	7498
571	Qy	GlnIleArgSerAspArgValAsnProLysThrPheTyrAlaLeu- - - - -SerAsnGly	588
7499	Db	TTTCCCTGCATCTCATCTGTAAATTCGAGTAAATTTTATCATCTTCGATTGGAAAAACGGT	7558
589	Qy	ThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeuPro	608
7559	Db	GACTTCCTCATAGTAAGATGAGGAAAGAGTATTATGAAA- - -GGTGCTAAACTTCCT	7615
609	Qy	SerSerGlyAlaValGlyValMetPheHisAla- - - - -ValProGlyLysGlu	624
7616	Db	TCCTTTGATAACGTGGGTATCGCTTATTCTTCTCTCTGTTGGCTCTCGACAGAGAA	7675
625	Qy	GlyAspLeuTrpLeuAlaAlaSer- - -SerGlyLeuTyrHisSerThrAsnGlyGlySer	643
7676	Db	GGAGATATCTGGTTAGCATTCAATGAATGGACTTTACAGATCAAAAGATGTTGGAATA	7735

	Qy	644 SerTrpSerAlaIleThrGlyValSerSerAlaValasnValcglyPheGlyLysSerAla	663
	Dd	7736 ACCTTTTCAGCGCATGGAAACGTATTATAGCCTAGCTGATTGTTTTTGCAGGCCCAAAA	7795
	Qy	664 ProGlySerSeryrProAlavalPheValValcglyThrllecglyGlyValThrGlyValA	683
	Dd	7796 CCAAGTGACTCATACTCACGAANTTAACCTGAACGGTATGCTGTAATGGAGTTTATGCTATT	7855
	Qy	684 TyrArgSerAspAspCysgLythrThrTriprValLeulleasnaAspaspGlnHisGlnTry	703
	Dd	7856 TTCATGTCCACAGATCAGGCGCAAACAATGATGAGAATTAACAACGATATAGCATCAATTT	7915
	Qy	704 GlyAsmTripglyGlnalalieThrGlyVspHisaalaenleuargvalvTyrieGly	723
	Dd	7916 GGA---TGGATACATTACATGATAGAGATATAGAAATTCGGAGAAATTTTCCTTCGT	7972
	Qy	724 ThrAsnGlyArgglylleValtyrGlyVspille	734
	Dd	7973 ACCGAAGGAAGCAATCATCGTTGTGTGAAGTG	8005

Search completed: May 12, 2004, 09:25:37  
Job time : 11278.8 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 12, 2004, 03:07:10 ; Search time 1026.48 Seconds

(without alignments)  
3062.577 Million cell updates/sec

Title: US-09-917-376-3

Perfect score: 4036

Sequence: 1 ATTQPTWSNVAIGGGFVD.....YIGTNGRGIVYDIGGAPSG 740

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09917376/runat\_11052004\_114534\_28231/app\_query.fasta\_1.11429  
-DB=N Geneseq 29Jan04 -QFWT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFWT=ptp -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09917376@cgn\_1\_1758@runat\_11052004\_114534\_28231 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N Geneseq\_29Jan04:\*  
1: Geneseqn1980s:\*  
2: Geneseqn1990s:\*  
3: Geneseqn2000s:\*  
4: Geneseqn2001as:\*  
5: Geneseqn2001bs:\*  
6: Geneseqn2002s:\*  
7: Geneseqn2003as:\*  
8: Geneseqn2003bs:\*  
9: Geneseqn2003cs:\*  
10: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4036	100.0	2869	7 ABZ77632	Abz77632 Nucleotid
2	4036	100.0	2869	9 ADD22922	Add22922 Acidothe
3	2063	51.1	2823	7 ABV76941	Abv76941 Nucleotid
4	1506	37.3	2849	9 ADD42060	Add42060 Trichoder
5	1139	28.2	3668	9 ADD24893	Add24893 DNA encod
6	878.5	21.8	5698	9 ADD42054	Add42054 Trichoder
7	392.5	9.7	1103	3 AAF14988	Aaf14988 Trichoder
8	238	5.9	7407	7 ACA42281	Aca42281 Prokaryot

9	238	5.9	7407	8 ACC59398	Acc59398 Microbial
10	199.5	4.9	7305	7 ACA23399	Aca23399 Prokaryot
c 11	195	4.8	5295	7 ACA24001	Aca24001 Prokaryot
c 12	192	4.8	9903	7 ACA40315	ACA40315 Prokaryot
c 13	192	4.8	110000	4 AAI99682_04	Continuation (5 of
c 14	184	4.6	3324	7 ACA38418	Aca38418 Prokaryot
c 15	184	4.6	110000	4 AAI99683_04	Continuation (5 of
c 16	182	4.5	11148	7 ACF93934	AcF93934 Mycobacte
c 17	182	4.5	110000	4 AAI99682_37	Continuation (38 o
c 18	182	4.5	110000	4 AAI99683_37	Continuation (38 o
c 19	181.5	4.5	11679	8 ADA30090	Ada30090 DNA encod
c 20	179.5	4.4	3222	7 ACA26562	Aca26562 Prokaryot
c 21	177.5	4.4	110000	4 AAI99682_30	Continuation (31 o
c 22	177.5	4.4	110000	4 AAI99683_30	Continuation (31 o
c 23	176	4.4	32367	2 AAV35620	Aav35620 Human SHO
c 24	174.5	4.3	67459	7 ABQ77491	Abq77491 S. aurant
c 25	174	4.3	2855	4 ABL29005	AbI29005 Drosophil
c 26	174	4.3	5221	4 ABL19628	AbI19628 Drosophil
c 27	174	4.3	5302	4 ABL29004	AbI29004 Drosophil
c 28	173	4.3	2708	4 ABL19629	AbI19629 Drosophil
c 29	170.5	4.2	2176	4 ABL11445	AbI11445 Drosophil
c 30	170.5	4.2	4252	4 ABL11444	AbI11444 Drosophil
c 31	170.5	4.2	7191	7 ACA47178	Aca47178 Prokaryot
c 32	170.5	4.2	25165	8 ACD19026	Acd19026 E. coli 0
c 33	170	4.2	4323	7 ACA20513	Aca20513 Prokaryot
c 34	170	4.2	110000	4 AAI99683_37	Continuation (38 o
c 35	169.5	4.2	3360	9 AAD62307	Aad62307 Candida a
c 36	169	4.2	110000	4 AAI99682_37	Continuation (38 o
c 37	169	4.2	110000	4 AAI99682_38	Continuation (39 o
c 38	168.5	4.2	25473	9 ADC01011	Adc01011 Enteroha
c 39	168	4.2	5451	6 ABZ35486	Abz35486 Human gen
c 40	166	4.1	3879	7 ACA20847	Aca20847 Prokaryot
c 41	165.5	4.1	4383	9 AAD62312	Aad62312 Candida a
c 42	165	4.1	6615	7 ACA40310	Aca40310 Prokaryot
c 43	165	4.1	110000	4 AAI99682_03	Continuation (4 of
c 44	165	4.1	110000	4 AAI99683_03	Continuation (4 of
c 45	164.5	4.1	3237	7 ABT42169	Abt42169 Toxicity

#### ALIGNMENTS

RESULT 1

ABZ77632

ID ABZ77632 standard; DNA; 2869 BP.

XX ABZ77632;

XX ABZ77632;

DT 03-JUN-2003 (first entry)

XX Nucleotide sequence of the avicelase AvIII.

XX Avicelase; AvIII; glycoside hydrolase; enzyme; cellulase; biofuel;  
XX detergent; pulp processing; paper processing; feed processing; textile;  
XX cellulose; gene; ss.

XX Acidothermus cellulolyticus.

XX Key Location/Qualifiers

PH CDS 1..2869

FT /\*tag= a

FT /partial

FT /product= "AvIII"

FT /transl\_except= (pos:2869 aa:Xaa)

FT /note= "Xaa is an unspecified residue"

XX WO2003012090-A2.

XX 13-FEB-2003.

XX 28-JUL-2001; 2001WO-US023818.

XX 28-JUL-2001; 2001WO-US023818.

(MIDE ) MIDWEST RES INST.

Ding S, Adney WS, Vinzant TB, Himmel ME;

WPI; 2003-248177/24.

P-PSDB; ABP73015.

New thermostable AviiiI peptide from *Acidothermus cellulolyticus*, useful for degradation of cellulose or in generating anti-AviiiI antibodies for purifying recombinant AviiiI polypeptides from genetically engineered host cells.

Claim 3; Page 24; 44pp; English.

The present sequence encodes a thermostable avicelase polypeptide, designated AviiiI. AviiiI is a member of the glycoside hydrolase family of enzymes, and is a cellulase. AviiiI is useful in the conversion of biomass to biofuels and biofuel additives. It may be useful in the production of detergents, pulp and paper processing, food and feed processing and in textile processes. The thermostable AviiiI peptide is useful in the degradation of cellulose, and in generating specific anti-AviiiI antibodies that are useful in purifying recombinant AviiiI polypeptides from genetically engineered host cells, in detecting AviiiI polypeptide expression, as well as a reagent tool for characterizing the molecular actions of the polypeptide. The AviiiI polynucleotide is useful as a source of probes or primers in various diagnostic assays

SQ Sequence 2869 BP; 546 A; 857 C; 900 G; 565 T; 0 U; 1 Other;

## Alignment Scores:

Pred. No.: 9,08e-252 Length: 2869  
Score: 4036.00 Matches: 740  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 7 Gaps: 0

US-09-917-376-3 (1-740) x ABZ77632 (1-2869)

PA	XX	1	AlaThrThrGlnProTyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAsp	20
PI	XX	139	GCACGACTCGCGTACACTCGGACGACGTCGCGATCGGGCGCGGCTTGTGAC	198
DR	XX	21	GlyLeuValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGly	40
DR	XX	199	GGGATCGTCTCAATGAAGGTGCACCGGGAATCTGTACGTGCGGACGGACATCGGGGG	258
XX	XX	41	MetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuAspTrpValGlyTrp	60
PT	XX	259	ATGTATCGATGGGATGCGGCAACGGCGCGTGGATCCCTCTCTGGATTGGTGGATGG	318
PT	XX	61	AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys	80
PS	XX	319	AACAATGGGGGTACACGCGCGTCTCAGCATTCGGCGACCGCGATCAATACTAA	378
XX	XX	81	ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu	100
XX	XX	379	GTATGGCCCGCTCGGAATGTACACCAACAGCTGGGACCCCAACGAGCGGATCTC	438
PT	XX	101	ArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGly	120
XX	XX	439	CGCTCGTCTCATCGGCGCGCAAGTGGCAATAAAGCCCTCGCGTCAAGCTTGGCGGC	498
XX	XX	121	AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle	140
XX	XX	499	AACATGCCCGGGCTGGATGGCGGAGCGGCTTGGCGTGGATCCAAACAATGACAA	558
XX	XX	141	LeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThr	160
XX	XX	559	CTGTATTTGGCGCCCGCGGCGCAAGAGGCTCTGGAGAAAGCACAGATTCGGCGCG	618
XX	XX	161	TrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThr	180
XX	XX	619	TGGTCCCGAGTACGAACTTTCCGGACGTAGGCACGTACATTGCAATCCCACTGACAG	678
XX	XX	181	ThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSer	200
XX	XX	679	ACCGCTATCAGAGCGATATTCAAGGCGTCTGTGGTCTGCTTTCGACAAGTCTTCTG	738
XX	XX	201	SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProVal	220
XX	XX	739	TCGTTCGGCGCAAGGAGTAGAACCAATTTTGTGGCGTGGCGGATCCCAATAATCC	798
XX	XX	221	PheTrpSerArgAspGlyValAlaThrTrpGlnAlaValProGlyValAlaProThrGly	240
XX	XX	799	TTCTGGACGAGACGGCGCGGACGTGGCGGCGTGGCGGCGCGCGCGCTTC	858
XX	XX	241	IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsn	260
XX	XX	859	ATCCCGCACAGGGCGTCTTTGACCCGCTCAACACAGTGTCTATATTGCCACCAAT	918
XX	XX	261	ThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGly	280
XX	XX	919	ACGGTGGTCTCGTATACGGGAGCTCCGGACGCTCGGAAATCTCGGTGACCTCCGG	978
XX	XX	281	ThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyr	300
XX	XX	979	ACATGGACCGCAATCAGCCCGTACTTTCGACGACGCGGCAACGACTACTTTGGTTAC	1038
XX	XX	301	SerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSer	320
XX	XX	1039	AGCGCTCTACTATCGACCGCGCAGCACCGGACAGATATGTTGGCAACCCAGATATCG	1098
XX	XX	321	TrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIle	340
XX	XX	1099	TGGTGGCGGACACCAATAATCTTTCGAGACCGGCGGTCGACGCGGAGCGCGATC	1158
XX	XX	341	TrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGlu	360
XX	XX	1159	TGGGATTGGACGAGTTATCCCAATCGAAGCTTGCATATGTCTTGACATTTCCGGGAG	1218
XX	XX	361	ProTrpLeuThrPheGlyValGlnProAsnProValProSerProLysLeuGlyTyr	380
XX	XX	1219	CCTTGGCTGACTTGGGCTACAGCGCAATCTCTCCGTACCCAGTCCGAGCTCGGCTGG	1278
XX	XX	381	MetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGly	400
XX	XX	1279	ATGGATGAAGCGATGGCAATCGATCCGTCAACTCTGTATCGGATGCTCTACGAA	1338
XX	XX	401	AlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIle	420
XX	XX	1339	CGGACGTTGTACGCAACAAATGATCTCAGAACTGGGACTCCGGCGGCGGAGATTCATATC	1398
XX	XX	421	AlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIleSerProSer	440
XX	XX	1399	CGCGCGATGGTCAAGGATTCGAGAGCGCGGTAAACGATCTCATAGCCCGCGCTCT	1458
XX	XX	441	GlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThr	460
XX	XX	1459	GGCGCCCGCTCATCGCGCTCTCGAGACCTCGCGCGCTTCCACCCCGCGGCTACT	1518
XX	XX	461	AlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyr	480
XX	XX	1519	GCCGTGCCATCGAGATCTTCACGTCACCGGTGTTCAGACCGGCGGACCGCTCGACTAT	1578
XX	XX	481	AlaGluLeuAsnProSerIleValArgAlaGlySerPheAspProSerSerGlnPro	500
XX	XX	1579	CGGAAATGAATTCGTTCGATCATCTTCGCGCTGGAACTTCATCCATCGAGCCCAACCG	1638
XX	XX	501	AsnAspArgHisValAlaPheSerThrAspGlyLysAsnTrpPheGlnGlySerGlu	520
XX	XX	1639	AACGACGACGCTCGCGTCTTCGACAGACGGCGGCAAGAACTGGTCCCAAGGACCGAA	1698
XX	XX	521	ProGlyGlyValThrThrGlyThrValAlaAlaSerAlaAspGlySerArgPheVal	540
XX	XX	1699	CCTGGCGGGTGACGACGGGCGGACCGCTCGCGCATCGGCGGACGGCTCTCGTTTCGTC	1758

QY 541 TrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrp 560  
 Db 1759 TGGGGTCCCGCGATCCCGGTGAGCTGGTGGTGTACGAGTCCGATTTGGCAACTCCTGG 1818  
 QY 561 AlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLys 580  
 Db 1819 GCTGCTTCGCAAGGTGTTCCCGCAATGCCAGATCCCGTTCAGACGGGTGAATCCAAAG 1878  
 QY 581 ThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPhe 600  
 Db 1879 ACTTTCATGCTATCCCAATGAACTTCTATCGAAGCAGCGCGCGGTGACATTC 1938  
 QY 601 GlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaVal 620  
 Db 1939 CAACCGGTCCGCGCGGTCTCCGAGCAGCGGTGCGGTGTCATGTTCCACGGGTG 1998  
 QY 621 ProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsn 640  
 Db 1999 CCTGGAAGAAAGGCGATCTGTGGCTCGCTGCATCGAGCGGGCTTTACCACCTCAACCAAT 2058  
 QY 641 GlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGly 660  
 Db 2059 GCGCGGAGCAGTGTGCTGCAATCAACCGCGGTATCTCCGCGGTGAACGTGGATTTGCT 2118  
 QY 661 LysSerAlaProGlySerSerTyrProAlaValPheValValGlyThrIleGlyGlyVal 680  
 Db 2119 AAGTCTGCGCCGCGGTGCTCATACCCAGCGCTTGTGTCGGCACGATCGAGGCGTT 2178  
 QY 681 ThrGlyAlaTyrArgSerAspCysGlyThrThrTrpValLeuIleAsnAspGln 700  
 Db 2179 ACGGGGGGTACCGCTCCGACGACTGTGGGACGACCTGGGTACTGATCAATGATGACCAG 2238  
 QY 701 HisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgVal 720  
 Db 2239 CACCAATACGGAATTTGGGGACAAGCAATCACCGGTGACCCGCAATTTACGGGGGTG 2298  
 QY 721 TyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyGlyAlaProSerGly 740  
 Db 2299 TACATAGGCACGAACGCGCGTGAATGTATACGGGGACATTTGGTGGTGGCGGTCGCGA 2358  
 RESULT 2  
 ID ADD22922 standard; DNA; 2869 BP.  
 XX AC ADD22922;  
 XX DT 15-JAN-2004 (first entry)  
 XX DE Acidothermus cellulolyticus avicelase AvIII DNA.  
 XX KW AvIII; cellulose reduction; agricultural biomass; municipal solid waste;  
 KW glycoside hydrolase; avicelase; ds; gene.  
 XX OS Acidothermus cellulolyticus.  
 XX FH Key Location/Qualifiers  
 FT CDS 1..2868  
 FT /\*tag= a  
 FT /product= "AvIII"  
 XX PN US2003108988-A1.  
 XX PD 12-JUN-2003.  
 XX PF 18-OCT-2002; 2002US-00155400.  
 XX PR 28-JUL-2001; 2001US-00917376.  
 XX PA (DING/) DING S.  
 PA (ADNE/) ADNEY W S.  
 PA (VINZ/) VINZANT T B.  
 PA (HIMM/) HIMMEL M E.

XX Ding S, Adney WS, Vinzant TB, Himmel ME;  
 XX WPI; 2003-810853/76.  
 DR P-PSDB; ADD22921.  
 XX New isolated thermal tolerant avicelase polynucleotide useful for  
 PT detection of a polynucleotide encoding AvIII and for reducing cellulose  
 PT in a starting material, e.g. municipal solid waste.  
 XX Claim 17; SEQ ID NO 2; 29pp; English.  
 XX The invention relates to an isolated polynucleotide molecule encoding a  
 CC thermostable AvIII polypeptide. The polynucleotide is useful for  
 CC detection of a polynucleotide encoding AvIII. The polynucleotide is  
 CC useful for reducing cellulose in a starting material which involves  
 CC administering to the starting material, e.g. agricultural biomass or  
 CC municipal solid waste, a polypeptide molecule of the polynucleotide. The  
 CC method further comprises administering a second polypeptide molecule  
 CC chosen from the glycoside hydrolase family of proteins. The present  
 CC sequence represents DNA encoding the Acidothermus cellulolyticus  
 CC avicelase AvIII.  
 XX Sequence 2869 BP; 546 A; 857 C; 900 G; 565 T; 0 U; 1 Other;  
 SQ  
 Alignment Scores:  
 Pred. No.: 9,08e-252 Length: 2869  
 Score: 4036.00 Matches: 740  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: Gaps: 0  
 US-09-917-376-3 (1-740) x ADD22922 (1-2869)  
 QY 1 AlaThrThrGlnProTyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAsp 20  
 Db 139 GCGACGACTCAGCGGTACACTCGAGCAACGTGGCGATCGGGCGCGCTTGTGTCAC 198  
 QY 21 GlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGly 40  
 Db 199 GGGATCGTCTTCAATGAAGGTGCACCGGGAATTCGTACGTGCGGAGCATCGGGGG 258  
 QY 41 MetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrp 60  
 Db 259 ATGTATCGATGGGATGCCGCCAACCGGCGGTGGATCCCTCTTCGGATTGGTGGGATGG 318  
 QY 61 AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 80  
 Db 319 AACAAATTGGGGTACAAACGCGTCTGACGATTCGGCGACACCCGATCAATTAACAAG 378  
 QY 81 ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 100  
 Db 379 GTATGGCGCGCGTCCGGAATGTACACCAACAGCTGGGACCCAAACGACGAGCGATTCTC 438  
 QY 101 ArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGly 120  
 Db 439 CGCTGCTGATCAGGGCGCAACGTGGCAATAAAGCCCTTCGCGTTCACGCTTGGCGGCG 498  
 QY 121 AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle 140  
 Db 499 AACATGCCCGGGGTGGATGGCGAGCGGCTTGGGGTGGATCCAAACATGACACAATT 558  
 QY 141 LeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThr 160  
 Db 559 CTGTATTTCCGGCGCCCGAGCGGCAAGGGCTCTGGAGAAGCACAGATTCCGGCGCGACC 618  
 QY 161 TrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThr 180  
 Db 619 TGGTCCAGATGACGAACCTTTCGGGACGTAGGACGCTACATTGCAAAATCCCACTGACACG 678  
 QY 181 ThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSer 200



DR WPI; 2003-092855/08.  
 DR P-PSDB; ABB99489.  
 XX New xyloglucanase enzyme belonging to family 74 of glycosyl hydrolases,  
 PT and endogenous to a bacterium, useful in the textile industry for  
 PT improving properties of cellulosic fibers, yarn or fabric.  
 XX Claim 11; Page 68-72; 76pp; English.  
 XX The present sequence encodes a xyloglucanase enzyme, belonging to family  
 CC 74 of glycosyl hydrolases. The enzyme is isolated from *Jonesia* sp.  
 CC DSM14140. The enzyme is useful in processes for machine treatment of  
 CC fabrics. It is also useful in the textile industry for improving the  
 CC properties of cellulosic fibers, yarn, woven or non-woven fabric, and in  
 CC a textile scouring process step. The xyloglucanase enzyme is also useful  
 CC in the cellulose fiber processing industry for rattling of fibers such as  
 CC hemp, jute, flax and linen. It is useful for preventing binding of  
 CC certain soils to the xyloglucan left on the cellulosic material  
 XX  
 SQ Sequence 2823 BP; 665 A; 803 C; 764 G; 591 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 3,88e-124 Length: 2823  
 Score: 2063.00 Matches: 386  
 Percent Similarity: 67.21% Conservative: 112  
 Best Local Similarity: 52.09% Mismatches: 225  
 Query Match: 51.11% Indels: 18  
 DB: 7 Gaps: 9

US-09-917-376-3 (1-740) x ABV76941 (1-2823)

QY 6 TyrThrSerAsnValAlaIleGlyGlyPheValAspGlyIleValPheAsn 25  
 DB 130 TATAGCTGTGACAGCTTCAGATTGTGGGCGGGTTTGTGCTGGATTGTCTTTAAC 189  
 QY 26 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrpAsp 45  
 DB 190 CAGAAAGACCCAGGCTGTGTGTATGCCCGCACAGATATTGGTGGAGCTACGGCTCAAT 249  
 QY 46 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr 65  
 DB 250 GACTCCACCGGGGTGGATCCCTCTTACGGACCATATTGGTTGGGATGACTGGAGCCAC 309  
 QY 66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 85  
 DB 310 TCCGGAATCTGTCTGTGGCCACTGACCTGTGTGACACCAACCGCTGTACCTTGGACGA 369  
 QY 86 GlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerAspGln 105  
 DB 370 GGCACGTACTCCATGGACTGGGATCCACAAACGGTGTCTATTGCTTCCGGCGGACAA 429  
 QY 106 GlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArg 125  
 DB 430 GGCAGACCTGGGAAAGACGATGCTGCGGCTCCGTGTGGGAGGCAACATGCGCGGTGCG 489  
 QY 126 GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAla 145  
 DB 490 GGGATGGGAGACGCTCCCGTGTGACCCCAACAATAAATGAAGTCTGTACTTTGTGTGG 549  
 QY 146 ProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThr 165  
 DB 550 GAATCGGGTATGAGCATATGGAAGTCACTACTACGGGAAACCTGGGGCAAGTGCAGC 609  
 QY 166 AsnProAspValGlyThrTrpIleAlaAsnProThrAspThrGly---TyrGln 184  
 DB 610 AGCTTCCCAATGCTGGTAACCTACGTCGCG-----GATGCTTGTGTGGGTACACC 660  
 QY 185 SerAspIleGlnClyValValTrpValAlaPheAspLysSerSerSerSerLeuGlyGln 204  
 DB 661 GGGCAAAACCAAGGTTGTGGTGGTAACTTTTGACCCCAACATCAGCAAGAGCTGTGTCAG 720  
 QY 205 AlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSerArg 224

DB 721 ACAACACAAACGATCTACGTAGGTGTAGCAGACAGCAGAAACAGTTTATCGATCTACC 780  
 QY 225 AspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIleProHisLys 244  
 DB 781 GATGGCGGGCAACATGGCAACGCGTCCCTGGCAACCCACCGGATTCTCGCGCAAAAG 840  
 QY 245 GlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyPro 264  
 DB 841 GGAGTGTCTTACCATAAAGGTACAGCAACTCTACATCGCTACCTCAGATATCGTGGCGCG 900  
 QY 265 TyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrpArg 284  
 DB 901 TATGACGGGTCGAAGGGGATGTGGCGCTAGATATTCCAGTGGCCAGTGGACTCGA 960  
 QY 285 IleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThr 304  
 DB 961 ATCAGCCCAATTCCTCTACC---TCATCGAATTCAGGTTCCGGTACTCCGTTTAGCC 1017  
 QY 305 IleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrpProAsp 324  
 DB 1018 ATTGACCGCAAAACCTGACACCATCATGGTTGTTTCCCAAGTGTCTTGTGGCTGAC 1077  
 QY 325 ThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIleTrpAspTrpThr 344  
 DB 1078 ATGTACGTATATCGTCCACCGACCGGGGGAACACGTGTCGCTATCTGGAACTGAAC 1137  
 QY 345 SerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpIleThr 364  
 DB 1138 GGTTCACAGCGCGGTACAAAGCAGTACACCAAGATTACTCAGGGGCTCCCTGGCTGAC 1197  
 QY 365 PheGlyValGlnProAsnProValProSerProLysLeuGlyTrpMetAspGluAla 384  
 DB 1198 TTTGGCAATCAGCCAAAGAGCCTGAAGCTAAACCCCAAACTCGGTGATGATGCCAGTCC 1257  
 QY 385 MetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyr 404  
 DB 1258 TTCGAGATTACCGCGCATATAACTCCGACCGCTTCTTATGTCACCGCGCGGCTAT 1317  
 QY 405 AlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetVal 424  
 DB 1318 GCGGAAACGAACCTCAAAACTGGGATAAGGGCAAAAGTTGATATCACTGTCAAGGCC 1377  
 QY 425 LysGlyLeuGluThrAlaValAlaValAsnAspLeuIleSerProProSerGlyAlaProLeu 444  
 DB 1378 CAAGGATCGAAGAAACCGCAGCCCAAGACTTCCCGCTCCCTCGGAAACATTTGGCCTA 1437  
 QY 445 IleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaValProSer 464  
 DB 1438 TACTCAGCGCTGGCAGACATCGGTGGCTTACCCCAAAAGATATCTCTCAGGTGCTAAC 1497  
 QY 465 Thr---IlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeu 483  
 DB 1498 AAGTACTACTACAAGAACCCACCATGACACCGTACCTCATTTGACTTGGCCGAAGC 1557  
 QY 484 AsnProSerIleIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArg 503  
 DB 1558 AAACCGCAACTGTGCTGCTGAGGCAAA-----TCTATTTCTGGGGAAACCACTCA 1611  
 QY 504 HisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGly 523  
 DB 1612 TGGTTGGGTTTCAACTGATGCTGGCGAAACGTTGAAACCTGGCGCCACCGCATCAGGG 1671  
 QY 524 ValThrThrGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPro 543  
 DB 1672 GTTAAAGCCCGGCTCCATCATGTGTGCTAGCTAATGCTTCACTCATTTGTGTGGCGCGG 1731  
 QY 544 GlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSer 563  
 DB 1732 -----GAAGTGGCGCTCTCGAGCTTCCACCAACAGTGGCTCTCTGTGTCAAGGTG 1785  
 QY 564 GlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyr 583  
 DB 1786 AGCGGGTCTCTCACAATGCCAGGTGCGCATCAGACCGAGTCAACGCTAATACTCTTTTAC 1845

Qy 584 AlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyValThrPheGlnProVal 603  
 Db 1846 GGCTTCGTTACCGGAAGTTCATCACTCCACAAATGGTGGGCTTCCTTCACCGCGTCA 1905  
 Qy 604 Ala---AlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaValProGly 622  
 Db 1906 GCTTTCACGGGGTTCACCAAGTGA-----AACGTTGTTTCCTGCGCGTCCGGGC 1959  
 Qy 623 LysGluGlyAspLeuTrpLeuAla-----AlaSerSerGlyLeuTyrHis 637  
 Db 1960 CGTCAAGTTCATCTCTGCTTGTGTTGGATATCGGGAAGCACTATGGATGTGGCG 2019  
 Qy 638 SerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnVal 657  
 Db 2020 TCCACTGATCGCGGAAGAACTGACCAAGTATCTCGAGTTTCAGGAAGCGATGCCGTT 2079  
 Qy 658 GlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValValGlyThrIle 677  
 Db 2080 GGTTCGTAAGCAACAGTAGTAGCGGATACCACTGATTTTCATCCGCGCAAGATT 2139  
 Qy 678 GlyGlyValThrGlyAlaTyrArgSerAspCysGlyThrThrTrpValLeuIleAsn 697  
 Db 2140 GATGGATTTCAGGAATTTTCGGTCTGACATGAGCAAGACCTGGAACGCATCAAC 2199  
 Qy 698 AspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeu 717  
 Db 2200 GACAACCAACACAGTGTATGACTGGCGCATCAATCACCAGTGCACCCAGATGTCGAA 2259  
 Qy 718 ArgArgValTyrIleGlyThrAsnGlyArgGlyValThrGlyAspIleGlyGlyAla 737  
 Db 2260 GGGCGGCTTACATTGGTACCAACGGTGGCGGATATCGTTGGTGTGATTCCTCCACTCCC 2319  
 Qy 738 Pro 738  
 Db 2320 CCG 2322  
 RESULT 4  
 ADD42060  
 ID ADD42060 standard; DNA; 2849 BP.  
 XX  
 AC ADD42060;  
 XX  
 DT 15-JAN-2004 (first entry)  
 XX  
 DE Trichoderma reesei Family 74 xyloglucanase gene, SEQ ID NO:1 #2.  
 XX  
 KW Trichoderma reesei; Family 74 xyloglucanase; beta-1,4-glycosidic linkage;  
 KW xyloglucan hydrolysis; biomass degradation; cellulose; hemicellulose;  
 KW ethanol production; detergent composition; fabric treatment;  
 KW textile treatment; gene; ds.  
 XX  
 OS Hypocrea jecorina.  
 XX  
 FH Key Location/Qualifiers  
 CDS 59..2743  
 FT /\*tag= b  
 FT /product= "Family 74 xyloglucanase"  
 FT /note= "Contains 3 introns"  
 FT 59..381  
 FT /\*tag= a  
 FT /number= 1  
 FT 382..440  
 FT /\*tag= c  
 FT /number= 1  
 FT 441..1227  
 FT /\*tag= d  
 FT /number= 2  
 FT 1228..1277  
 FT /\*tag= e  
 FT /number= 2  
 FT 1278..2664  
 FT /\*tag= f

FT intron  
 FT 2665..2723  
 FT /\*tag= g  
 FT /number= 3  
 FT 2724..2743  
 FT /\*tag= h  
 FT /number= 4  
 XX  
 FN WO2003089598-A2.  
 XX  
 XX 30-OCT-2003.  
 XX  
 PF 17-APR-2003; 2003WO-US011831.  
 XX  
 PR 19-APR-2002; 2002US-0373987P.  
 XX  
 PA (NOVO ) NOVOZYMES BIOTECH INC.  
 XX  
 XX Michael R, Zaretsky E, Haas J;  
 PI WPI; 2003-845528/78.  
 XX  
 DR P-PSDB; ADD42055.  
 XX  
 XX New polypeptides having Family 74 xyloglucanase activity, and encoding  
 FT nucleic acid molecules, useful for degrading cellulose- and hemicellulose  
 PT -containing biomass to ethanol or as a detergent.  
 XX  
 XX Claim 1; Fig 1; 96pp; English.  
 PS  
 CC The invention relates to a Family 74 xyloglucanase (ADD42055) from the  
 CC fungus Trichoderma reesei (Hypocrea jecorina), and nucleic acids encoding  
 CC it (e.g., ADD42054). The enzyme hydrolyses the beta-1,4-glycosidic  
 CC linkages in the backbone of xyloglucan to xyloglucan oligosaccharides.  
 CC The invention also relates polypeptide sequences at least 70% identical  
 CC to the enzyme, expression vectors and host cells comprising a nucleic  
 CC acid of the invention, the recombinant production of the enzyme, and  
 CC mutant enzymes and the nucleic acids encoding them. The xyloglucanase of  
 CC the invention can be used in the degradation of cellulose- and  
 CC hemicellulose-containing biomass to produce ethanol. It can also be used  
 CC in a detergent composition for treating fabric during a machine washing  
 CC cycle. The present sequence represents the gene encoding Trichoderma  
 CC reesei Family 74 xyloglucanase. Note: A comparison of this sequence with  
 CC that shown in the sequence listing (also referred to as SEQ ID NO:1)  
 CC shows that the present sequence is half the length of the sequence  
 CC listing SEQ ID NO:1 (ADD42054), as the sequence listing SEQ ID NO:1  
 CC contains alternate 60 bp blocks of sense strand and complementary strand  
 CC sequence.  
 XX  
 SQ Sequence 2849 BP; 594 A; 882 C; 822 G; 551 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 4,23e-88 Length: 2849  
 Score: 1506.00 Matches: 329  
 Percent Similarity: 56.60% Conservative: 121  
 Best Local Similarity: 41.38% Mismatches: 260  
 Query Match: 37.31% Indels: 87  
 DB: 9 Gaps: 23  
 US-09-917-376-3 (1-740) x ADD42060 (1-2849)

Qy 6 TyrThrTrpSerAsnValAlaIle---GlyGlyGlyGlyPheValAspGlyIleValPhe 24  
 Db 119 TTTTCATGGAAGACGTCACGTCGGCGCGCGCTTCGTCGCCGCGCATCTTC 178  
 Qy 25 AsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrp 44  
 Db 179 CATCCCAAGACAAAGGCGGTAGCATATGCACCAAGATATTGGCGGCTGTACCGCTC 238  
 Qy 45 AspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpVal-----GlyTrp 60  
 Db 239 AAC---GCCGACGACTCATGACCGCGCTGACGGATGGATTCGTATATATCCGCGCTG 295  
 Qy 61 AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 80





## RESULT 5

ADD24893  
ID ADD24893 standard; DNA; 3668 BP.

XX AC ADD24893;

XX DT 15-JAN-2004 (first entry)

XX DNA encoding *Xanthomonas campestris* cellulase #1.

XX Directed genetic engineering; galactomannanase; reduced activity;  
XX enhanced activity; xanthan gum production; suspension stability;  
XX emulsion stability; temperature resistance; pseudoplasticity; amylose;  
XX cellulase; extracellular protease; intracellular protease;  
XX glucose dehydrogenase; gene; ds.

XX *Xanthomonas campestris*.

XX US2003036176-A1.

XX 20-FEB-2003.

XX 10-AUG-2001; 2001US-00927827.

XX 28-MAR-2001; 2001US-0279493P.

XX (BOWE/) BOWER S G.

XX (RAMS/) RAMSEIER T M.

XX Bower SG, Ramseier TM;

XX WPI; 2003-625389/59.

XX P-PSDB; ADD24919.

XX New transformed cell or organism having reduced or enhanced activity of  
XX at least one protein, useful for producing xanthan gum, which are useful  
XX for providing formulations and properties.

XX Claim 15; SEQ ID NO 21; 135pp; English.

XX The present invention relates to polypeptide and polynucleotide sequences  
XX from *Xanthomonas campestris* which may be used for activity reduction or  
XX enhancement using directed genetic engineering. A transformed cell or  
XX organism having reduced or enhanced activity of at least one such protein  
XX e.g. galactomannanase can be generated by disrupting the gene encoding  
XX the protein. The activity of the protein is reduced by the presence of an  
XX antisense nucleic acid sequence. The nucleic acid sequence of the gene  
XX encoding the protein is a recombinant sequence having at least one  
XX mutation as compared to the wild-type gene encoding the protein. The  
XX transgenic cell or microorganism are useful for producing xanthan gum,  
XX which are useful for providing formulations and properties, such as long-  
XX term suspension and emulsion stability in alkaline, acid, and salt  
XX solutions, temperature resistance, and pseudoplasticity. The present  
XX sequence encodes an enzyme relating to the present invention.

XX SQ Sequence 3668 BP; 570 A; 1181 C; 1292 G; 625 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 3.13e-64 Length: 3668  
Score: 1139.00 Matches: 267  
Percent Similarity: 50.85% Conservative: 120  
Best Local Similarity: 35.09% Mismatches: 295  
Query Match: 28.22% Indels: 79  
DB: 9 Gaps: 21

US-09-917-376-3 (1-740) x ADD24893 (1-3668)

Qy 1 AlaThrGlnProTyrThrTrpSerAsnValAlaTleGlyGlyGlyPheValAsp 20

Db 1217 GCCACGTCGGGGCCCTACCAAGTCGCGAGTGGCCATTCGGCGTGGCGCTTGTGACC 1276

Qy 21 GlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGly 40

Db 1277 GGTGTGCTGTTTCATCCCGCAACGTCGTATGCGCTATGCGCCACCGCATGTGGGTGGC 1336  
Qy 41 MetTyrArgTrpAspAlaAalaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrp 60  
Db 1337 GCGTACCGCTGGGATGCGCAGCGCAGTGGACCGCGCTGACGACGTGGTGGCGCT 1396  
Qy 61 AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 80  
Db 1397 GACGACTGGAACTGATGGGCATCGACGATTCGCGTCGACCCCGCGCATCCGATCG 1456  
Qy 81 ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 100  
Db 1457 CTGTATCTGGGCGCGGCACCTATATGATGAA---CGCGCGGCACCTGCGCGAGTGTG 1513  
Qy 101 ArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGly 120  
Db 1514 CGCTCGTTCAACCGCGCGCGCACGTTGAGCGTGCAGCTTCCGCTTTAAGTGGGTGGT 1573  
Qy 121 AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle 140  
Db 1574 AACGAGCTGGCGCGCGCAATGGCGAGCGGTGGCGTGGACCCGACGATGGCGCGGTG 1633  
Qy 141 LeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThr 160  
Db 1634 CTGCTGCTGGGCTCGCGCGATGCC---GGCTGTGGCGTAGCGAGTATCGCGCGCGCAC 1690  
Qy 161 TrpSerGlnMetThrAsnProAspValGlyThrTyrIleAlaAsnProThrAspThr 180  
Db 1691 TGGCGGAAGTGGCGTGTTCGCGACCGCGCGTGGCGGTGCCACCGCGCGCAATCAT 1750  
Qy 181 ThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSer 200  
Db 1751 GTTGGCGCGAGCAGCGGTG---GGATCGCTTTGTGTCGTGTCGACGACCGACGATG 1807  
Qy 201 SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProVal 220  
Db 1808 AACACCGCTCGCAACACCGCGCATCTACGTGGCGGTGCCACCGAACACGACGACCTG 1867  
Qy 221 PheTrpSerArgAspGlyGlyValaThrTrpGlnAlaValProGlyValaProThrPhe 240  
Db 1868 TATGTGTCGGAAGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 1927  
Qy 241 IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsn 260  
Db 1928 CGCCCGAGCCACATGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1984  
Qy 261 ThrGlyGlyProTyrAspGlySerGlyAspValTrpLysPheSerValThrSerGly 280  
Db 1985 CAGCCCG 2044  
Qy 281 ThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyr 300  
Db 2045 CGCTGGCGTGAGATCAGCCCGATTCGCGAG---CCAGCCAGTGGCGATGATGCTGCTG 2101  
Qy 301 SerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSer 320  
Db 2102 GGTGCGGTGGCGGTGGATCCGCAACATCCGAGGTGCTGCTGGCGCGCGCGCGCGCT 2161  
Qy 321 TrpTrpProAspThrIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIle 340  
Db 2162 CGCAGCG 2221  
Qy 341 TrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGlu 360  
Db 2222 -----TTGCCCGATGCGGTGTTTCGATTCACAGCCCGCGCG 2254  
Qy 361 ProTrpLeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGlyTrp 380  
Db 2255 CCGTGG---ACCGCACATGCCACCGCGCAC-----TGG 2284  
Qy 381 MetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGly 400

Db 2285 ATGGGG--CGCGTGGGATCGATCGTTCCAGCGGCAACCATCGGCTGTTCTGTCGACCGGC 2341  
 Qy 401 AlaThrLeuTyxAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIle 420  
 Db 2342 TACGGCATCTGGGCTCGGCAATCTG-----CAGGATTTTCGCC 2380  
 Qy 421 AlaPro-----MetValIysGlyLeuGluGluThrAlaVal 432  
 Db 2381 GCACCGCAGCGCCGCTGCAGTGGTGTTCAGGACCGTGGCTGGAAGAAACCGTCCCG 2440  
 Qy 433 AsnAspLeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGly 452  
 Db 2441 CTGAGCTCTCAGCCCGATGGCTGGCGGCATCTGCTACGCGCTCGCGCATATCGAC 2500  
 Qy 453 GlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPhe 472  
 Db 2501 GGCTTCGGCATCAGCACCTCGACCGGTG--CAGTTGAGTACCGCGGCCACCGGTG 2557  
 Qy 473 ThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGly 492  
 Db 2558 ACCAATGGCGAAAGCATCATGATCGCGCGCGCAGCGCGCGTGGTGGTGGCGAGCGGT 2617  
 Qy 493 SerPheAspProSerSerClnProAsnAspArgHisValAlaPheSerThrAspGlyGI 512  
 Db 2618 ACCGTGCGCGCGCGCGCGCAACAAATCCGCGCGC---TGATTACCGGTGATGGCGG 2674  
 Qy 512 YLysAsnTrpPheGln---GlySerGluProGlyGlyValThrThrGlyGlyThrValAl 531  
 Db 2675 CAGCATGAGTGGCTGTTGCGAGTGAGCGCCAGCGCGCGCGCGCGCGCATTCG 2734  
 Qy 531 aAlaSerAlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValVa 551  
 Db 2735 CATTTGGTCCGATCGCGCGCAGGTGGTGTGGCAGCGCGCGCGCGCGCAAT-----TG 2788  
 Qy 551 lTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGI 571  
 Db 2789 GCGCAGCTCCGACTTCGGCGCGCAGTGGCAGCGTGTGAACGCCCTGCGCAACCCCGCGT 2848  
 Qy 571 nIleArgSerAspArgValAsnProLysThrPheTyrAla-----LeuSerAsnGlyTh 589  
 Db 2849 GGTGTGACCGATCGGTGGTGGCAGCAGCGCGCTGGTATGCGGTGATCGCCAGCGGCA 2908  
 Qy 589 rPheTyrArgSerThrAspGlyGlyValThrPheGlnProValAlaAlaLeu----- 607  
 Db 2909 GCTGTACGAGACCGATCCCGCGCGCAGTTCCTG-----GCGACCGGTGTACAGGT 2962  
 Qy 608 -----ProSerSerGlyAlaVa 613  
 Db 2963 GGGCAGCGCGCGCGATGAGCGCACCGCGCGCAGCTGCTCCGACCCGTCGGCGCGC 3022  
 Qy 613 lGlyValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSe 633  
 Db 3023 AGCGGTGGTGTATCTGGCGACCGCGCGCAAG----- 3053  
 Qy 633 rGlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSe 653  
 Db 3054 -GGTGTATCGCTGGCAGACCGT-----GCCTCAGGTGCTCTCAGACCGCGGACGA 3106  
 Qy 653 rAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheVa 673  
 Db 3107 AGCAGCTCGCTGGCATCGGAAGGCGTGGCGCGCGTGGCGCGCGGTGTATCT 3166  
 Qy 673 lValGlyThrIleGlyValThrGlyAlaTyrArgSerAspAspCysGlyThrThrTr 693  
 Db 3167 GCGCGCGCGCTGACGAGCGGTGGTGTGCTTCGCTCCGACGATGGCGCGCGCAGTG 3226  
 Qy 693 pValLeuIleAsnAspAspGlnHisGlnIntyrglyAsnTrpGlyGlnAlaIleThrGlyAs 713  
 Db 3227 GCAGCGCATCAACATGACCGCGCACCGCTTCGGCGCGC---CCGTACAGCGTACCGGTGA 3283  
 Qy 713 pHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAs 733  
 Db 3284 TCCGCGCATTCGCGGGGTGTGTACTTCGCGCACCGCGCGCGCGCATTTTCTACGCGCA 3343

Qy 733 p 733  
 Db 3344 T 3344

## RESULT 6

ADD42054  
 ID ADD42054 standard; DNA; 5698 BP.

AC ADD42054;

DT 15-JAN-2004 (first entry)

XX Trichoderma reesei Family 74 xyloglucanase-encoding DNA, SEQ ID NO:1 #1.  
 XX Trichoderma reesei; Family 74 xyloglucanase; beta-1,4-glycosidic linkage;  
 XX xyloglucan hydrolysis; biomass degradation; cellulose; hemicellulose;  
 XX ethanol production; detergent composition; fabric treatment;  
 XX textile treatment; ds.

OS Hypocrea jecorina.

XX WO2003089598-A2.

XX 30-OCT-2003.

XX 17-APR-2003; 2003WO-US011831.

XX 19-APR-2002; 2002US-0373987P.

XX (NOVO ) NOVOPYMES BIOTECH INC.

XX Michael R, Zaretsky E, Haas J;

XX WPI; 2003-845528/78.

XX P-PSDB; ADD42055.

New polypeptides having Family 74 xyloglucanase activity, and encoding  
 nucleic acid molecules, useful for degrading cellulose- and hemicellulose  
 -containing biomass to ethanol or as a detergent.

Claim 1; SEQ ID NO 1; 96pp; English.

The invention relates to a Family 74 xyloglucanase (ADD42055) from the  
 fungus *Trichoderma reesei* (*Hypocrea jecorina*), and nucleic acids encoding  
 it (e.g., ADD42054). The enzyme hydrolyses the beta-1,4-glycosidic  
 linkages in the backbone of xyloglucan to xyloglucan oligosaccharides.  
 The invention also relates polypeptide sequences at least 70% identical  
 to the enzyme, expression vectors and host cells comprising a nucleic  
 acid of the invention, the recombinant production of the enzyme, and  
 mutant enzymes and the nucleic acids encoding them. The xyloglucanase of  
 the invention can be used in the degradation of cellulose- and  
 hemicellulose-containing biomass to produce ethanol. It can also be used  
 in a detergent composition for treating fabric during a machine washing  
 cycle. The present sequence represents a *Trichoderma reesei* Family 74  
 xyloglucanase-encoding DNA. Note: The present sequence (referred to as  
 SEQ ID NO:1) is given as shown in the sequence listing. A comparison of  
 this sequence with that shown in figure 1 (also referred to as SEQ ID  
 NO:1) shows that the present sequence is twice as long as the figure 1  
 sequence (ADD42060) and contains alternate 60 bp blocks of sense strand  
 and complementary strand sequence.

Sequence 5698 BP; 1145 A; 1704 C; 1704 G; 1145 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.:	3.69e-47	Length:	5698
Score:	878.50	Matches:	334
Percent Similarity:	29.92%	Conservative:	119
Best Local Similarity:	22.06%	Mismatches:	270
Query Match:	21.77%	Indels:	793
DB:	9	Gaps:	52

US-09-917-376-3 (1-740) x ADD42054 (1-5698)



```
Db 2400 CACTGAGACCTATTACTACAGCATCTCAGTGAAGTCACTCTTAACGATCCGATCGGAATGG 2459
QY 363 LeuThrPhe-----GlyValGlnProAsn----- 370
Db 2460 CTGACTCTGGATAATGATGTCGTAGAGTCACCTCAGTGCAGAAATGCTAGGCTAGCTTACC 2519
QY 371 -----ProProValPro----- 374
Db 2520 GACTGACTCTGCTTCAAGACTCCCAAAGCACCGTGGATCAAGAAACAACCTTTATCGATGTG 2579
QY 374 ----- 374
Db 2580 ATGACTGAGACGAAGTTCTCAGGGTTTCGTGGCACCCTAGTCTCTTGTGAAATAGCTTACAC 2639
QY 375 -----SerPro-----LysLeuGlyTrpMetAspGluAlaMet 385
Db 2640 TCGAGCGAGTCACCGTCCGATGCTCTCAACAGCGCCTCGGCTGGATGATTGAGTCTCTC 2699
QY 386 Ala----- 386
Db 2700 GGCTCGCTCAGTGGCAGGCTACAGAGTAGTTTCGGCGGCGGACCTACTAACTCAGAGAG 2759
QY 387 ---IleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAla 405
Db 2760 CAGATTGACCAACCGACAGCAACCACTGGCTCTACGGCACCGGAATGACAATCTTTGGC 2819
QY 406 Thr----- 406
Db 2820 GTCTAACTGGGTTGGCTGTGTTGTGTGACCGAGATCCGTGGCCTTACTGTTTGAAGAACCG 2879
QY 407 ---AsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValLys 425
Db 2880 CGCCAGATCTCACCAACTGGGACACGGCCACAAATGTGTAATCCAATCACTGGCAGAC 2939
QY 426 Gly----- 426
Db 2940 GCGGTCTAGAGTGGTTGACCTGTGCGCGGTGTTACACAGTTAGTTAGTGACCGCTCTG 2999
QY 427 ---LeuGluGluThrAlaValAsnAspLeuIleSerProSerGlyAlaProLeuIle 445
Db 3000 CGCATCGAGGAATTTCTCCGTCCAGGACCTGGCTCTGCAACCCGCGGAGCGAGCTATTG 3059
QY 446 Ser----- 446
Db 3060 GCGTAGCTCTTAAGAGGAGGTCTGACCGGAGAGCGTGGCGCGCTTCGCTCGATAAC 3119
QY 447 ---AlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaVal----- 462
Db 3120 CCGCGAGTCGAGACGACAAACGCTTCACTTTGCCAGCAGAAACGACCTCGGGACATCG 3179
QY 462 ----- 462
Db 3180 CGGCGTCAGCTCTGCTGTTGCCAAGTGGAAACGGTCGCTTTGCTGGAGCCCTGTAGC 3239
QY 463 ProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAla--- 481
Db 3240 GCGCAGACGCTCTGGGCAACGCCCCACATGGGCCACCTCGACAGCGCTCGACTACGCGGG 3299
QY 482 -----Glu 482
Db 3300 AGGCTCTGCCAGACCGTTGCGGGGTGTACCGGTGGAGTGTCTCGCAGCTGATCGGCCCC 3359
QY 483 LeuAsnProSerIleValArgAlaGlySer----- 493
Db 3360 TACTCGGTCAAGAGCTGCTGCGCGTCCGCAACACACCGCGGACGCAACAGGTGGCCATC 3419
QY 494 -----PheAspProSerSerGlnProAsnAspArgHisValAlaPhe----- 507
Db 3420 TTGAGCCAGTTCTTCGACGAGCGGCGGCGGCTTGTGGCGCGCGTGTGTCACCGGTAG 3479
QY 508 SerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGlyValThrGly 527
Db 3480 ACGTCCGACGGCGGCGGACGCTGGAGCATCGACTACGGGCCGACACGTCATGAACGCG 3539
QY 528 Gly----- 528
Db 3540 GGCAGGCTGCGCGCGCTGCACCTCTAGTGTGATGCGCGGCTGTGAGGTACTGTGCG 3599
QY 529 ---ThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaProGlyAspProGly 547
Db 3600 CGCACGCTGGCCTATTTCGGCGGACGGCGACACGATCCTCTGCTCGACCGCCTCTGTCGCGC 3659
QY 548 -----GlnPro 549
Db 3660 GCGTCCACCGGATAACCGGCTGCCGCTGTGTAGAGACCAGCTGGCGGAGCAGGCG 3719
QY 550 ValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsn 569
Db 3720 CTCGACGCTCG---CAGTTCCAGGCGAGCTTTTGCCTCCGTCCTCGAGCTTGCCTCGGCG 3776
QY 570 AlaGln----- 571
Db 3777 GCGGACGTCGCGAGCGTCAAGGTCCTCGTCAAAACGGAGCGACAGCTCGGACGGGCGCGC 3836
QY 572 ---IleArgSerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThr 589
Db 3837 CGGCTCATCGCTCGGACAAAGAACCAACAGCGCTCTTACCGCGGCTCGGATCGACC 3896
QY 590 Phe----- 590
Db 3897 TTTTAGTAGCGGACCTGTTCTTCTGTTGTCGAGAGATGCGGCGGAGGCTAGCTGG 3956
QY 591 ---TyrArgSerThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSer 609
Db 3957 AAAAACTGACGAAGGACACCGGCGAGCAGCTTC-----ACGCGCGGG---CCCAAG 4004
QY 610 SerGlyAlaValGlyValMet----- 616
Db 4005 CTGGGACGCGGAGGATGCGATCGTCTGTGGCGCTGTCGAAAGTGGCGCGCGCGGTTTC 4064
QY 617 -----PheHisAlaValProGlyLysGluGlyAspLeu 627
Db 4065 GACCCGTGCGGTCCTCGATCCGGGATATCGCTGCTACCCGACACCGCGGCGGACGTTG 4124
QY 628 TrpLeuAlaAlaSerSer----- 633
Db 4125 TATGCTCGACCGCGGCTAGGCCCTATAGCAGAGTGGCTGTGTGGCGCGCGCTGCAAC 4184
QY 634 -----GlyLeuTyrHisSerThrAsnGlyGlySerSer----- 644
Db 4185 ATACAGAGCTGGCTGCTCGGCATATTTCGCTCCACAGACTCGGGCAGCACCTTTGGCCAA 4244
QY 644 ----- 644
Db 4245 GTCTCCACCGCCCTGAACCGGTATGAAGCGAGGTGTCTGAGCCCGCTGTGGAACCGGTT 4304
QY 645 ---TrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSer 662
Db 4305 CAGAGTGGCGGACTCCAAACACCTACAGATCGCCTGGGTGTGGCTCAGCTCGAAC 4364
QY 663 -----AlaProGlySerSerTyr 668
Db 4365 TGGAACTGTATGCTGTTGTGATGTTCTAGCGGAGCCACACACCGAGTCCGAGCTTG 4424
QY 669 ProAlaValPheValValGlyThrIleGlyGlyValThrGlyAla-----TyrArgSer 686
Db 4425 ACCTTGGACATACGATCGGCACC-----GGCCCGTCAGGGGCTCGCCTCAGCCAGT 4478
QY 687 AspAspCysGlyThrThrTrpVal----- 694
Db 4479 GGAGACAGCGCGCTCTCTGAAGCCGTGGCGGCGAGTCCCGGAGCGGAGATGCGGTCA 4538
QY 695 -----LeuIleAsnAspAspGlnHisGlnTyrGly----- 704
Db 4539 CCTCTGCGCGGAGGAGACCTCGACATCCAGGGCTCCAGGGGCTTCGGCTCCATCGAC 4598
```

Qy 704 ----- 704  
 Db 4599 AGCACCAAGTCCCGCAGCGGCTGTAGTCCCGAGGGTCCCGAAGCCGAGGTAGCTG 4658  
 Qy 705 AsnTyrGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgArgValTyrIleGlyThr 724  
 Db 4659 TCGTGGTTCAGCGCGCGTCCGCGAGCACC CGCGG-----CAAGTCTACGTGGGCACC 4712  
 Qy 725 AsnGlyArgGlyIleValTyrGlyAspIleGlyAlaPro 738  
 Db 4713 AACGGCCGGCGGCTTTTACGCTCAGCGCTCGTGGCGGCC 4754

RESULT 7  
 AAF14988  
 ID AAF14988 standard; cDNA; 1103 BP.  
 XX  
 AC AAF14988;  
 XX  
 DT 15-SEP-2003 (revised)  
 DT 13-MAR-2001 (first entry)  
 XX  
 Trichoderma reesei EST SEQ ID NO:7511.  
 XX  
 Multiple gene expression; filamentous fungal cell; EST;  
 KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;  
 KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;  
 KW culture condition; environmental stress; spore morphogenesis;  
 KW metabolic pathway engineering; catabolic pathway engineering; ss.  
 XX  
 OS Hypocrea jecorina.  
 XX  
 WO200056762-A2.  
 XX  
 28-SEP-2000.  
 XX  
 22-MAR-2000; 2000WO-US007781.  
 XX  
 22-MAR-1999; 99US-00273623.  
 XX  
 (NOVO ) NOVO NORDISK BIOTECH INC.  
 PA (NOVO ) NOVO NORDISK AS.  
 XX  
 Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;  
 PI WPI; 2000-594572/56.  
 XX  
 Monitoring differential expression of genes in filamentous fungal cells  
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a  
 PT substrate of expressed sequence tags.  
 XX  
 Claim 89; Page 3034; 3161pp; English.  
 PS  
 XX  
 CC The present invention describes a method for monitoring differential  
 CC expression of genes in a first filamentous fungal (FF) cell relative to  
 CC expression of the same genes in one or more second filamentous fungal  
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from  
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs  
 CC are used in the methods for monitoring differential expression of genes  
 CC in a first filamentous fungal (FF) cell relative to expression of the  
 CC same genes in one or more second filamentous fungal cells. Monitoring the  
 CC global expression of genes from FF cells allows the production potential  
 CC of the microorganisms to be improved. New genes may be discovered.  
 CC possible functions of unknown open reading frames can be identified and  
 CC gene copy number variation and stability can be monitored. The expression  
 CC of genes can be used to study how FF cells adapt to changes in culture  
 CC conditions, environmental stress, spore morphogenesis, recombination,  
 CC metabolic or catabolic pathway engineering. Using ESTs provides several  
 CC advantages over genomic or random cDNA clones including elimination of  
 CC redundancy as one spot on an array equals one gene or open reading frame,  
 CC and organization of the microarrays based on function of the gene  
 CC products to facilitate analysis of the results. AAF07478 to AAF11247  
 CC represents ESTs from Fusarium venenatum; AAF11248 to AAF11853 represents

CC ESTs from Aspergillus niger; AAF11854 to AAF14878 represents ESTs from  
 CC Aspergillus oryzae; and AAF14879 to AAF15337 represents ESTs from  
 CC Trichoderma reesei, which are all specifically claimed in the present  
 CC invention. (Updated on 15-SEP-2003 to standardise OS field)  
 XX  
 SQ Sequence 1103 BP; 230 A; 322 C; 307 G; 223 T; 0 U; 21 Other;

## Alignment Scores:

Pred. No.: 1.54e-16 Length: 1103  
 Score: 392.50 Matches: 114  
 Percent Similarity: 45.05% Conservative: 36  
 Best Local Similarity: 34.23% Mismatches: 125  
 Query Match: 9.72% Indels: 58  
 DB: 3 Gaps: 10

US-09-917-376-3 (1-740) x AAF14988 (1-1103)

Qy 302 GlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTyr 321  
 Db 3 GGCCTTGGCCTCGATTTGCAAAAGCCAGGACCCTTGTGTGCTTCTTTGAACCTCTGG 62  
 Qy 322 TrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTyrThrArgIleTyr 341  
 Db 63 TGGCAGATGCTCAGCTGTTTCGGTGCACCGACTCTGGGACCAACATGAGCCGATCTGG 122  
 Qy 342 AspTyrThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluPro 361  
 Db 123 CGGTGGCGGAGCTATCCGACTGAGACCTATTACTACAGCATCTCAACTCCCAAGCACCG 182  
 Qy 362 TrpLeuThrPheGly---ValGlnProAsnProProValProSer-----Pro 376  
 Db 183 TGGATCAAGAACAACTTTTATCGATGTGACGCGAGTCCATCCGTCGATGCTCTNATCAAG 242  
 Qy 377 LysLeuGlyTyrMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeu 396  
 Db 243 CGCTCGGCTCGATGATTGAGTCTNCTGAGATTGACCCACCGCAGCAANNACTGGCTT 302  
 Qy 397 -TyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrIleTyrAspSer----- 414  
 Db 303 TTACGGCACCGGAATGACAAATTTTGGCGGCCACGATTTTCCAACTGGGACACGGCCCC 362  
 Qy 415 -----GlyGlyGlnIleHisIleAlaProMetValLys-GlyLeuGluGluThrAlaV 432  
 Db 363 ACAATGTGGTCAATCCCAATTACTTGGCAGACGGGATTGGAAGGAATT-----TTTCCGT 416  
 Qy 432 alaAsnAspLeuIleSerProProSerGly----- 441  
 Db 417 TCAAGGACCTGGCGCTTTTACCCCGGGGGAAGCGAGCTTTTGGCCGCAAGTCCGGAGACG 476  
 Qy 442 -----AlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspV 459  
 Db 477 ANCAACGGGTTTACCTTTTGGCCAGCAGAAACGACCTTGGG----- 517  
 Qy 459 alThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThr-SerVal 478  
 Db 518 -----ACATTCGCCGACAGCGGTTTGGGCACTCCACATGGGCCACCTCGAGAAAGCGTC 572  
 Qy 479 AspTyrAla--GluLeuAsnProSerIleIleValArgAlaGlySer----- 493  
 Db 573 GACTACCGCGGAATACTCGTCAAGAGCGCGTTCGTCCGCGTCGCGCAACACCGCCGCGCAG 632  
 Qy 494 -----PheAspProSerSerGlnProAsnAspArgHisValAlaPheS 508  
 Db 633 CAACAAGGTGGCCATTTTTCGAACGGCGGGCGCCGACGTNGAA-----GCAATTC 686  
 Qy 508 erThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGlyValThrThrGlyG 528  
 Db 687 GAACATAAGCT-----GGTCCGAACACCGTTTCCATTGCAACAGCGG 728  
 Qy 528 lyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaProGlyAspProGlyG 548  
 Db 729 GCGCGGTGGCGCTATTTCGCGCCGACGACGACGATCTCTGTGTCGACCGCGCTCGTCCGGC- 787

QY 548 lnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProA 568  
 Db 788 -----GTGAGCGCTCGAGTTCAGGCGGAGCTTCCTCCGTCGAGCTGCCCC 839  
 QY 568 laAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnG 588  
 Db 840 CGGGCGCGTGCATCGNCTGGCAAGAACACACGAGCGTTCCTAGCGCGCTCGGAT 899  
 QY 588 lYThrPheTyrArgSerThrAspGlyGly 597  
 Db 900 CGACCTTTAGTTCAGTTCAGCAAGGACACCGGC 928

## RESULT 8

ACA42281

ID ACA42281 standard; DNA; 7407 BP.

XX AC ACA42281;

XX DT 19-JUN-2003 (first entry)

XX DE Prokaryotic essential gene #23938.

XX KW Antisense; ds; prokaryotic essential gene; cell proliferation;

XX KW drug design; gene.

XX OS Pseudomonas aeruginosa.

XX PN WO200277183-A2.

XX PD 03-OCT-2002.

XX PF 21-MAR-2002; 2002WO-US009107.

XX PR 21-MAR-2001; 2001US-00815242.

XX PR 06-SEP-2001; 2001US-00948993.

XX PR 25-OCT-2001; 2001US-0342923P.

XX PR 08-FEB-2002; 2002US-00072851.

XX PR 06-MAR-2002; 2002US-0362699P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KJ, Zyskind JW;

XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WPI: 2003-029926/02.

XX DR P-PSDB; ABUS8411.

XX New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX Claim 14; SEQ ID NO 30151; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of

CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
 CC prokaryotic essential genes. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 7407 BP; 1387 A; 2763 C; 2378 G; 879 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 1.36e-05 Length: 7407  
 Score: 238.00 Matches: 205  
 Percent Similarity: 32.59% Conservative: 88  
 Best Local Similarity: 22.80% Mismatches: 314  
 Query Match: 5.90% Indels: 292  
 DB: 7 Gaps: 50

US-09-917-376-3 (1-740) x ACA42281 (1-7407)

QY 11 ValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGly 30  
 Db 3250 GTCAACCTGAGCAATGGCAGCAGCTCAGCGC-----ACCGCGAGCGCGC 3297  
 QY 31 IleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrpAspAlaAsnGlyArg 50  
 Db 3298 AGCAGCGTGTATCTCACCAGC-----GGCAACGCGCAAT 3330  
 QY 51 TrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr----- 65  
 Db 3331 CCGATCCCGAGGTACCGCGCAGCGAGCGGCACTGAGCACTACACCCGTCACGCGG 3390  
 QY 66 -----AsnGly---ValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrp 82  
 Db 3391 ATCGCCAAACGGTACTGTGGTCAACGTCGCGGCCAGGAGCGCTCCGGTAACAGCAGCCG 3450  
 QY 83 AlaAlaValGlyMetTyrThrAsnSer-----TrpAspProAsnAsp 96  
 Db 3451 CCGCGCAGCGGTGACCGTCCGATTCACGCGCGCGCGCGCGCGGTGATCAACCCGAGCAAC 3510  
 QY 97 GlyAlaIleLeuArg---SerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuPro 115  
 Db 3511 GCGTGTGTATCAGCGCAGCGCAGCGCGCGCGCGCGCGGTGACCGTCCCTCACC----- 3561  
 QY 116 PheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspPro 135  
 Db 3562 ---GATCCGCGCGGCAACCCGATAGGCGAGTACCGCGCGAC----- 3600  
 QY 136 AsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThr 155  
 Db 3601 -----GGCAGCGGCAACTGGAGCTTCACG 3624  
 QY 156 AspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAla 175  
 Db 3625 CCG-----GGCACCCTCGTGGCC----- 3642  
 QY 176 Asn-----ProThrAspThrThrGlyTyrGlnSer 185  
 Db 3643 AACGCAACGCTGATCTGCGCCACCGCCACCGCGGCAATACCGCGCGCGAGGCC 3702  
 QY 186 -----AspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerSer 201  
 Db 3703 GCCACCAACGCTGGAGCGGTGGCGCGCGCGCGGTGATCGATCCGAGCAACGCGCAG 3762  
 QY 202 Leu-----GlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsn---Asn 218  
 Db 3763 ACCATCAGCGGCAACCGCGAGCGCGGCGCAAGGTGATCTCTCACCAGCGGCAACGCGCAG 3822  
 QY 219 ProVal---PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaPro 237

```

Db 3823 CCGATCGCGGAACCAACCGCGGACGGGCAACTGGAGCTTTCACGCCCGGCAACGCCG 3882
Qy 238 -----|||||-----|||||-----ThrGly--- 239
Db 3883 CTGGCCAAACGCGCGGTGTAACCGCGTGGCCCGAGGACCTCGGGCAATACCGGCCG 3942
Qy 240 -----PheIleProHisLysGlyValPheAspProValAsn 251
Db 3943 CAGGGCAGCACTACCGTGGAGCGCGGTGGCGCGGAACACCGCTGTGGTCAATCCGACAA 4002
Qy 252 HisValLeuTyrIleAlaThrSerAsnThrClyGlyProTyr-----AspGlySer 268
Db 4003 GGCAACCTGTCTCAACGGTACCGCGCGAGCGCGGAGCACCGTGACCTTGACCGCGCAAC 4062
Qy 269 SerGlyAspValTrpLysPheSerValThr---SerGlyThrTrp-----ThrArgile 285
Db 4063 GGCAACCCGATCGCGCAGACCAACCGCGGATGGCAGCGGCACTGGAGTTCACGCCCGCG 4122
Qy 286 SerProValProSer-----ThrAspThrAlaAsnAspTyrPheGlyTyrSer 301
Db 4123 TCGCAACTGCGCCCAACGCGCACCGGTGTCACCGTACCGCGAGCGCGCGGCAATACC 4182
Qy 302 GlyLeu-----ThrIleAspArgGlnHisProAsnThrIleMetVal----- 315
Db 4183 AGCCTTCCCGCTACCGACGCGGTGGATTCTCGCTGCGGTGATCCCGCGAGTGGATCCG 4242
Qy 316 -----AlaThrGlnIleSerTrpTrpProAsp-----AlaThrTrpThrArgileTrpAsp 342
Db 4243 AGCAACCGTTCGGTGATCAGCGGCGCACCGCGGACCGCGGCAACACATCATCATC- 4296
Qy 331 ThrAspGlyGly-----AlaThrTrpThrArgileTrpAsp 342
Db 4297 ACCGATGGCAACGCAACCGGATTGGCGAGTCCACCGCGGCGGCAACCTGGTCC 4356
Qy 343 TrpThr-----SerTyrProAsnArgSerLeuArgTyrVal 354
Db 4357 TTCACTCCAGGCAFCGCGTGGCGGATGGCAGCGTGTCAACGTGGTGGCGCGGCGCA 4416
Qy 355 LeuAspIleSerAlaGluProTrpLeu-----ThrPheGlyValGlnProAsnProPro 372
Db 4417 AGCAATGTGCAGACGCGCGCGGTGATCACTGTGGTGGCGGTGGCGCGCGCGCGCG 4476
Qy 373 ValProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSer 392
Db 4477 GTG-----ATCGATCCGAGCAACGCG 4497
Qy 393 AspArgMetLeuTyrGlyThr-----GlyAlaThrLeuTyrAlaThrAsnAspLeu 409
Db 4498 ACCGAG---ATAAGCGGTACCGCGGAGCGCGCGGCGGCGGTCGTCCTCACC- 4545
Qy 410 ThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGlu 429
Db 4546 -----GATGCGCGCGGCAAC-----CCGATCGCGCGGCGGCGGCGGCGGCG 4584
Qy 430 ThrAlaValAsnAspLeuIleSerPro-----ProSerGlyAlaProLeuIle 445
Db 4585 GGCAAGCGCACTGGAGCTTCAACCGCGAGCACCCCGGTGGCCACGCGGACCGTGCATCAAC 4644
Qy 446 SerAlaLeuGlyAspLeuGlyGlyPheThrHis-----AlaAspVal 459
Db 4645 GCCGTGGCCAGGACCGCGCGGCGGCAATACCAACGCGGTCCGCGGCGGCGGCGGCGG 4704
Qy 460 ThrAlaValProSerThrIlePhe-----ThrSerProValPheThrGlyThrSer 477
Db 4705 ATCCCGCGCGCGCGCGCGGTGATCAATCCGAGCAATGAGTGCATCAGCGGTACG--- 4761
Qy 478 ValAspTyrAlaGluLeuAsnProSerIleValArgAlaGlySerPheAspProSer 497
Db 4762 -----GCGAGAGCGCGGCGGCGGCGGATCTCCAGGCGGCAACCGGCAACCCGATC 4812
Qy 498 SerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPhe--- 516

```

```

Db 4813 GCCCAG-----GTCACCGCGGACGGGCAACTGGAGCTTC 4851
Qy 516 ----- 516
Db 4852 ACGCCCGGCGCGCGCTGGCCCAACGGCTCGGTGATCAATGCGCTGGCCCGAGGACGCCG 4911
Qy 517 ---GlnGlySerGluProGlyGlyValThr----- 525
Db 4912 GGCAACACAGCAGCGTCCCAACGCGCCACCGTGCAGCTCGCTGGCGCCAGCAGCCCGGTG 4971
Qy 526 -----ThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrp 541
Db 4972 ATCGATCCGAGCAACGAGTAGCTGATCGCGGTACCGCGAGGTGGTGGTCCCGGTGATC 5031
Qy 542 AlaProGlyAspProGlyGlnProVal-----ValTyrAlaValAlaGlyPheGlyAsnSer 559
Db 5032 CTCACCGAGCGCAACGGCAACCGGATCGCGCAGGTCCACCGCGATGGCAGCGGCAAC--- 5088
Qy 560 TrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnPro 579
Db 5089 TGGAGCTTCACGCGCGCGGCAACCG- 5112
Qy 580 LysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArg----- 592
Db 5113 -----CTGTCCAATGGCACGGTGTCAATGCGGTGGCCCGAGGACGCTGCC 5157
Qy 593 ---SerThrAspGlyGlyVal-----ThrPheGlnProValAlaAlaGlyLeuPro--- 608
Db 5158 GGCAACACCGCGCGCGCTGAGCACCGGTGGAGCGGTGGCGCGCGCGCGCGCGCGGTG 5217
Qy 609 -----SerSerGlyAlaValGlyValMetPheHisAlaValProGlyLysGlyAsp 626
Db 5218 ATCGACCGCGCAACCGGTGTGAACTCAGCGCGCACCGCGCAACCGCGGTGCGGTGATC 5277
Qy 627 LeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSer---ThrAsnGlyGlySerSerTrp 645
Db 5278 CTCACCGATGCAATGGCAATCCGATCGCGCGCACGCTTGGCGCGGCGGCGGCAACTGG 5337
Qy 646 SerAlaIleThrGlyVal-----SerSerAlaValAsnValGlyPheGlyLys 661
Db 5338 AGCTTACCGCGCGGCGCGCGCTGGCCCAACCGCGCGGTGTCAATGCGGTGGCCCGAGGAC 5397
Qy 662 SerAlaProGlySerSerTyrProAlaValPheValValGlyThrIleGlyGlyValThr 681
Db 5398 CCGCGCGGCAATACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5457
Qy 682 GlyAlaTyrArgSerAspAspCysGlyThrThrTrpValLeuIleAsnAspAspGlnHis 701
Db 5458 -----CCGGTGTATCAATCCCGAGCAAC--- 5478
Qy 702 GlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgValTyr 721
Db 5479 -----GGCAGCGGTGATCACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5520
Qy 722 IleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyGlyAlaProSerGly 740
Db 5521 ATCTCTACCGATGGCAACGCGCAACCGGATCGCGGAGACCGCGCGCGCGCGCGCGCG 5577

```

## RESULT 9

ACC59398

ID ACC59398 standard; DNA; 7407 BP.

XX

AC ACC59398;

XX

DT 28-AUG-2003 (first entry)

XX

DE Microbial resistance gene PA1874 coding sequence.

XX

KW Biofilm; microbial resistance; gene; ds.

XX

OS Unidentified.

XX

PN WO2003041483-A2.



XX 22-MAY-2003.  
 XX  
 XX 18-SEP-2002; 2002WO-US029565.  
 XX PF  
 XX 18-SEP-2001; 2001US-0323241P.  
 XX PR  
 XX (DART-) DARTMOUTH COLLEGE.  
 XX PA  
 XX O'toole GA, Mah T;  
 XX PI  
 XX WPI; 2003-468567/44.  
 XX DR P-PSDB; ABP59933.  
 XX  
 XX Identifying modulators of microbial resistance of organisms in biofilms,  
 PT e.g. inhibitor of biofilm formation, by employing expression controls, or  
 PT efflux pumps containing polypeptides, of genes associated with biofilm  
 PT resistance.  
 XX  
 XX Disclosure; Fig 7; 102pp; English.  
 XX  
 CC The present invention relates to a method of identifying a compound  
 CC capable of altering the sensitivity of a microorganism to an  
 CC antimicrobial agent by employing efflux pumps comprising polypeptides  
 CC encoded by the following genes: PA1874, PA4142, PA2389, PA1876, PA4143,  
 CC PA2390 or PA1163. The method is useful for identifying modulators of  
 CC microbial resistance of an organism in a biofilm. The methods are also  
 CC useful for identifying genes that encode proteins that play a role in  
 CC biofilm resistance. The method is particularly useful for screening  
 CC compounds or discovering compositions that will inhibit biofilm formation  
 CC and overcome their resistance mechanisms. These methods are particularly  
 CC useful in medical, industrial or natural settings, where formation of  
 CC biofilms can have serious negative consequences and result in high costs  
 CC both in human health and economic terms. The present sequence is a  
 CC coding/control sequence shown in the exemplification of the invention  
 XX  
 XX Sequence 7407 BP; 1387 A; 2763 C; 2378 G; 879 T; 0 U; 0 Other;  
 XX  
 Alignment Scores:  
 Pred. No.: 1.36e-05 Length: 7407  
 Score: 238.00 Matches: 205  
 Percent Similarity: 32.59% Conservative: 88  
 Best Local Similarity: 22.80% Mismatches: 314  
 Query Match: 5.90% Indels: 292  
 DB: 8 Gaps: 50  
 US-09-917-376-3 (1-740) x ACC59398 (1-7407)  
 QY 11 ValAlaIleGlyGlyGlyPheValAspGlyValPheAsnGluGlyAlaProGly 30  
 DB 3250 GTCACTGAGCAATGGAGCAGCCTCAGCGGC-----ACCGCGAGCGCGGC 3297  
 QY 31 IleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArg 50  
 DB 3298 AGCAGCGTGATCTCACCGAC-----GGCAACGGCAAT 3330  
 QY 51 TrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr----- 65  
 DB 3331 CCGATCCGAGGTTCACCGCGCAGCGGCACTGGACCTACACCCCGTCCACCGCG 3390  
 QY 66 -----AsnGly--ValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrp 82  
 DB 3391 ATGCCCAACGGTACTGTGTGTCAACGTGTGGCCCGCAGGACGCTCCGGTTAACAGCAGCCG 3450  
 QY 83 AlaAlaValGlyMetTyrThrAsnSer-----TrpAspProAsnAsp 96  
 DB 3451 CCGCGCAGCGTGACCGTCGATTCCAGCGCGCGCGCGGTGTGATCAACCGCAGCAAC 3510  
 QY 97 GlyAlaIleLeuArg---SerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuPro 115  
 DB 3511 GCGGTGTCATCAGCGGCACCGCGGCGGTGTCCACCGTGCACCTCACC----- 3561  
 QY 116 PheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspPro 135

DB 3562 ---GATCCCGCGCAACCGCATAGGCGAGGTACCGCGAC----- 3600  
 QY 136 AsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThr 155  
 DB 3601 -----GGCAGCGGCAACTGGAGCTTCAGC 3624  
 QY 156 AspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAla 175  
 DB 3625 CCG-----GGCACCCCGCTGGCC 3642  
 QY 176 Asn-----ProThrAspThrThrGlyTyrGlnSer 185  
 DB 3643 AACGGCAGCGTGATCGTCGCCACCGCCACCGCAGCGCAATACCGCGCCCGCAGCGC 3702  
 QY 186 -----AspIleGlnGlyValValTrpValAlaPheAspLysSerSerSer 201  
 DB 3703 GCACACACGGTGGACGGGTGGCGCGCGCGGTGTGATCGATCCGAGCAACGGCAGC 3762  
 QY 202 Leu-----GlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsn---Asn 218  
 DB 3763 ACCATCAGCGGCACCGCGGAGCGCGGCAAGGTGATCTCACCGCAGCGCAACGGCAAC 3822  
 QY 219 ProVal---PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaPro 237  
 DB 3823 CCGATCGCGAAACACCGCGCGCAGCGCAACTGGAGCTTCACCGCGCAGCGCG 3882  
 QY 238 -----ThrGly--- 239  
 DB 3883 CTGGCCAAACGGCAGGTGGTCAACCGCGTGGCCCGAGGACCTTGGGGCAATACCGCGCG 3942  
 QY 240 -----PheIleProHisLysGlyValPheAspProValAsn 251  
 DB 3943 CAGGGCAGCACTACCGTGGACGGGTGGCGCGGCAACAGCCTGTGTCAATCCGAGCAAC 4002  
 QY 252 HisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyr-----AspGlySer 268  
 DB 4003 GGCAACCTGTCAACGGTACCGCGGAGCGCGGAGCACCGTGACCTTGACGCGCGCAAC 4062  
 QY 269 SerGlyAspValTrpLysPheSerValThr---SerGlyThrTrp-----ThrArgIle 285  
 DB 4063 GGCAACCGCATCGGCCACAGCACCGCGCGATGGCAGCGCACTGGAGCTTCACCGCGCGC 4122  
 QY 286 SerProValProSer-----ThrAspThrAlaAsnAspTyrPheGlyTyrSer 301  
 DB 4123 TCGCACTGCCCAACGGCAGCGGTGTCAACGTGACCGGAGCGCGCGCGCGCAATACC 4182  
 QY 302 GlyLeu-----ThrIleAspArgGlnHisProAsnThrIleMetVal----- 315  
 DB 4183 AGCTTCCCGTACCACGAGCGGTGATTCCTCGTGGCGTGCATCCCGCAGGTGGATCCG 4242  
 QY 316 -----AlaThrGlnIleSerTrpTrpProAsp-----ThrIleIlePheArgSer 330  
 DB 4243 AGCAACGGTTTCGGTGTATCAGCGGCACCGCGCAGCGCGCAACACCATCATCATC----- 4296  
 QY 331 ThrAspGlyGly-----AlaThrTrpThrArgIleTrpAsp 342  
 DB 4297 ACCGATGGCAACGGCAACCGGATGGCCAGGTACCGCGCGCAGCGCGCGCACTGGTCC 4356  
 QY 343 TrpThr-----SerTyrProAsnArgSerLeuArgTyrVal 354  
 DB 4357 TTCCTCCAGGCATCCCGTGGCGGATGGCAGCGGTGTCAACGTGTGGCGCGCGCAACCA 4416  
 QY 355 LeuAspIleSerAlaGluProTrpLeu-----ThrPheGlyValGlnProAsnProPro 372  
 DB 4417 AGCAATGTCAGACGTGCGCGCGGTGTATCACTGTGGATGGGTGGCGCGCGCGCGCGCG 4476  
 QY 373 ValProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSer 392  
 DB 4477 GTG-----ATCGATCCCGAGCAACGGC 4497  
 QY 393 AspArgMetLeuTyrGlyThr-----GlyAlaThrLeuTyrAlaThrAsnAspLeu 409



CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
 CC prokaryotic essential genes. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 QQ  
 XX Sequence 7305 BP; 1395 A; 2299 C; 2566 G; 1045 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 0.00414 Length: 7305  
 Score: 139.50 Matches: 185  
 Percent Similarity: 32.00% Conservatives: 72  
 Best Local Similarity: 23.04% Mismatches: 303  
 Query Match: 4.94% Indels: 243  
 DB: 7 Gaps: 42

US-09-917-376-3 (1-740) x ACA23399 (1-7305)

QY 47 AlaAsnGlyValTrp-----ilePro 53  
 DB 517 GCGACGGCGCTGTCGTACAGCCCGGTACGTCCTCCGAGGCGGCCCATCAGATACC 576  
 QY 54 LeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaA 73  
 DB 577 GTGACCGACAGCAGTAGTACGGGCATACGAGCGAGCTGAGCAGCGTGGGATCGTGTG 636  
 QY 74 AspProIle-----AsnThrAsnIysValTrpAla----- 83  
 DB 637 GATACGATCGTACAGAAACCGCTGCGATTACGCGCTCAGCGACGATGACAACTCCG 696  
 QY 84 --AlaValGlyMetTyrThrAsnSerTrpAspProAsn----- 95  
 DB 697 GTGACGATCGGCGGTGCGAGCAGCATACGACCGCGCTCGCGGTACCGGGAAGCG 756  
 QY 96 -----AspGlyValIleLeu-----ArgSerAspGln 105  
 DB 757 GGTAGCAGGCTGAGCTGTATGACGCGCAGCAGCTCTCTGGACGACGACGCGGATTCG 816  
 QY 106 GlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGly----- 119  
 DB 817 TCCGGGAAGTGGACATTTACGCGCACCG---GGTTGGCGAGCGCGCGCACTCGATT 873  
 QY 120 -----GlyAsnMetProGlyArgGlyMetGlyGluArgLeu 131  
 DB 874 ACGGTGACTGCGACGGATGCGCGCGCAACGTGAGCAGCAGCGCGGCTTCGAGCTG 933  
 QY 132 AlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeu 151  
 DB 934 AGATCGATACAGCGGCTCCGCGCTGCCCGTGAACGCGACGAGCGGTACGTGCTG 993  
 QY 152 TrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsn-----PhePro 168  
 DB 994 TCGGCGACCGGGAAGCGGCGGCGAGCTCAACATCGATACGATGGCGAGGTACGCG 1053  
 QY 169 AspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAsp----- 186  
 DB 1054 GAC---GCCACGCTGACGCGGTACCGGAGCGGCGTCTGGACGTACACGCGCTCGACCCCA 1110  
 QY 187 -----IleGlnGlyValValTrpVal---AlaPheAspLysSerSerSerLeuGly 203  
 DB 1111 CTGCCGATCGGACGGTGTATCGGCGTACGCGCGACGCGCGCGGCGGCAATACGGGCGCG 1170  
 QY 204 GlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSer 223  
 DB 1171 TCTGCTTCGTAACTGACGCGGCGGATACGCGGCGCGGTCGCGCGGTATC----- 1224  
 QY 224 ArgAspGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIleProHis 243  
 DB 1225 -----GGTACGCTGACGCGACGATGCG-----GGCTCTGTTGTTGCGATTGCTCT 1272

QY 244 LysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGly 263  
 DB 1273 GCGCGGAGTACTGATGATCGAGCGCCACATTTG---AGCGGCACAGCGGAAGCGGCGAGC 1329  
 QY 264 Pro-----TyrAspGlySer-----SerGly 270  
 DB 1330 ACAGTACGCTGTATGACGCGACGACCGTGTGGTACGACGACGCGACCGCCGTCCGGA 1389  
 QY 271 AspValTrpLysPheSerValThrSerGly-----ThrTrpThrArgIleSerProVal 288  
 DB 1390 AAC---TGGACGTTTACGCGCACCCAGCGGTTTGAAGAGCGCGCGACAGCCGTGACG 1446  
 QY 289 ProSerThrAspThrAlaAsnAspTyrPheGlyTyrSer-----GlyLeuThrIle 305  
 DB 1447 ACCGCCACGACACGCGCGCAATGTAGTGTTCGAGACGCGGCTTCGACCTGACGATC 1506  
 QY 306 AspArgGlnHisProAsnThrIleMetVal---AlaThrGlnIleSerTrpTrpProAsp 324  
 DB 1507 GATACGACAGCTCCGTTCGATTCGACAGTAGTAACCGGACCGAC-----GGT 1551  
 QY 325 ThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIleTrpAspTrpThr 344  
 DB 1552 ACCTCGCTGTCGGGACGCGGGAAGCGGCGCAACGATCAACATCGATACGACGCGGAC 1611  
 QY 345 SerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpIleThr 364  
 DB 1612 GGCACGCGCGATGCCAC-----GTAAACGCGACGACCCCGCGCGCG 1653  
 QY 365 PheGlyValGlnProAsnProValProSerProLysLeuGlyTrpMetAspGluAla 384  
 DB 1654 TGACGTFACGCGCTCGACCCCGCTCCGCGCT----- 1686  
 QY 385 MetAlaIleAspProPheAsnSerArgMetLeuTyrGlyThrGlyAlaThrLeuTyr 404  
 DB 1687 -----GSCACGCTGATTCGCGTGACG 1707  
 QY 405 AlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetVal 424  
 DB 1708 GCTACC-----GACGCGCGGCAATACGCGGCCCTCTCTCTTCGGA 1749  
 QY 425 LysGlyLeuGluThrAlaValAsnAspLeuIleSerProProSerGlyAlaProLeu 444  
 DB 1750 -----ACGCTGACGCGGATACGACGCGCGC-----GGTGGCGCGT 1788  
 QY 445 IleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaValProSer 464  
 DB 1789 ATCGTACGCTGACGAGCATGCGGCTCTCTGTTGGTGGATTCCTCTCGGCGCAT 1848  
 QY 465 ThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeuAsn 484  
 DB 1849 ACTGATGATCGCGCGCGACATTTG---AGCGGCACA-----GCGGACGACG 1893  
 QY 485 ProSerIleValArgAlaGly-----SerPheAspProSer 497  
 DB 1894 AGCAGGTGACGCTGTATACGCGACGACGCTTGTCTGGCAGCAGCAGCAGCGACCATCC 1953  
 QY 498 SerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPhe--- 516  
 DB 1954 GGA-----AACTGGACGTTT 1968  
 QY 517 -----GlnGlySerGluProGlyValThr-----Thr 526  
 DB 1969 AGCCCGACTACTGCTGGGTGAGGTGCACACAGCTGACCGTCACCGACGATACG 2028  
 QY 527 GlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaProGlyAspPro 546  
 DB 2029 GCGGCAACGTGAGTGTTCGAGCAGCGCTTCGATCTGACGATCGATACGACGCGCTCG 2088  
 QY 547 GlyGlnProValValTyrAlaVal---GlyPheGlyAsnSerTrpAlaAlaSerGlnGly 565  
 DB 2089 GCGATTCGACAGTGAACGCGACGAGCGGTACGCTGTTGCGGCGACGCGCGGCGGT 2148  
 QY 566 ValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrAlaLeu 585



Db 4844 GTCACATCGATACGAATGGCAGCGTACGCCGGAC---GCCACCGTACGGCTGACCCG 4788  
Qy 178 ThrAspThrThrGlyTyrGlnSerAsp-----IleGlnGlyValValTrpVal 193  
Db 4787 AGCGGGCTGTGACAGTACACGGCGTGCAGCCCACTGCCGATCGGCAGGTGATCGCGTG 4728  
Qy 194 ---AlaPheAspSerSerSerSerLeuGlyGlnAlaSerIleThrPheValGly 212  
Db 4727 ACGGCCACGCGCGCGGCAATACGGCCCGCTCTGCTTCGGTAACGGTACGGCGGAT 4668  
Qy 213 ValAlaAspProAsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAla 232  
Db 4667 AGACGGCACCGGGTCCCGGTCATC-----GGTACGGTACGACGATGCG 4620  
Qy 233 ValProGlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHis 252  
Db 4619 ---GGCTGTGTTGGTGGATGCTCTGGCGGCACTACTGATGATCGGACCGG 4566  
Qy 253 ValLeuTyrIleAlaThrSerAsnThrGlyGlyPro-----TyrAspGlySer--- 268  
Db 4565 ACATTG---AGCGGCACACGGAAGCGGCGCAGCAGCGTCAGCGTGTATGACGGCACGCG 4509  
Qy 269 -----SerGlyAspValTrpLysPheSerValThrSer 279  
Db 4508 TTGTTGGGTACACGACACCGCACCGTCCGGAAC---TGGACGTTACGCGCACCG 4452  
Qy 280 Gly-----ThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyr 297  
Db 4451 GGTGGGAAGCGGCGCAGACCTGACCGTCCACCGCACGACGCGCCGCGCATGTG 4392  
Qy 298 PheGlyTyrSer-----GlyLeuThrIleAspArgGlnHisProAsnThrIleMet 314  
Db 4391 AGTGTTCCGAGCAGCGCCTTCGACCTGACGATCAGCACGCTCGTCCGATCCGACA 4332  
Qy 315 Val---AlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAspGly 333  
Db 4331 GTGAACCGCACCGAC-----GGTACGTCGCTCGCGCACGCGCGGAAGCG 4287  
Qy 334 GlyAlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyr 353  
Db 4286 GCGCAACCATCAACATCGATACGAACGCGCGCAGCGCGCGGATGCCACC----- 4236  
Qy 354 ValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProProVal 373  
Db 4235 -----GTACACCGCACCGCGCGCGGTGGACGTACACGCGCTCGACCGCGCTG 4185  
Qy 374 ProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAsp 393  
Db 4184 CCGGCT----- 4179  
Qy 394 ArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrIleTrpAsp 413  
Db 4178 -----GGCACGGTGTATGGTGGCGGCTTACC-----GAC 4149  
Qy 414 SerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluThrAlaValAsn 433  
Db 4148 GCGCGGGCAATACGGCGCGCTCTGCTCGGTA-----ACGGTACGGCG 4104  
Qy 434 AspLeuIleSerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGly 453  
Db 4103 GATACGACGCGCG-----GGTGGCGCGTCACTCGGTACGTGACGACGATGCGGGC 4050  
Qy 454 PheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPheThr 473  
Db 4049 TCTGTTGTTGGTGGATGCTCTGGCGGCACTACTGATGATGCGACGCGCATTTG--- 3993  
Qy 474 ThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValargAlaGly--- 492  
Db 3992 AGCGGCACA-----GCGGAAGCAGGCGACGCGGTGAGCGGTATGACGCGCACG 3945  
Qy 493 -----SerPheAspProSerSerGlnProAsnAspArgHisValAla 506  
Db 3944 ACGTTGCTCGGCACGACGACGGGACCATCCGGA----- 3909

Qy 507 PheSerThrAspGlyGlyLysAsnTrpPhe-----GlnGly 518  
Db 3908 -----AATGACGTTACGCGCACTACTGCACTGGTGAGG 3870  
Qy 519 SerGluProGlyGlyValThr-----ThrGlyGlyThrValAlaAlaSerAlaAsp 535  
Db 3869 GCACACAGCTGACCGTCAACGCAACGATACGCGCGCAACGTGAGTGTCCGAGCAGC 3810  
Qy 536 GlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAlaVal--- 554  
Db 3809 GCCTTCGATCTGACGATCGATACGACGGCTCCGCGGATCCGACAGTGAACGCGACG 3750  
Qy 555 GlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSer 574  
Db 3749 GGTACGTGTTGTCGGCAGCGCCGAAAGCGGTCGACGGTCAACATCGATACGAATGCG 3690  
Qy 575 AspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThr 594  
Db 3689 GACGGCAGCGCGACGCCACCGTAACACGCCGACCCGACGCGCGCTGG- 3642  
Qy 595 AspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGly 614  
Db 3641 -----ACGTACACGCCC---TCGACCCCGCTCGCGCTGGTACGCTGATCGC 3597  
Qy 615 ValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGly 634  
Db 3596 GTGACGCCACCGATCGGCGACGCAAC----- 3570  
Qy 635 LeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAla 654  
Db 3569 -----ACGGGCGCGTCTGCTCGCTCG 3552  
Qy 655 ValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValVal 674  
Db 3551 GTAACGCTGACGGGTGACACGCGGACCGCGTCCGCG- 3507  
Qy 675 GlyThr-----IleGlyGlyValThrGlyAlaTyrArgSer- 686  
Db 3506 GGCACCGTACGATGACGTGGGTCTGTGTTGGTGGATCATTCGCGGCGAGTACT 3447  
Qy 687 AspAspCysGlyThrThrTrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrp 706  
Db 3446 GATGACGCGACGCGCATTTGAGC----- 3423  
Qy 707 GlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgArgValTyrIleGlyThrAsnGly 726  
Db 3422 GGTACGCGGAGGACGACGACGATGAGC-----GTGTATGACGCGACGACGCTG 3372  
Qy 727 ArgGlyIleValTyrGlyAspIleGlyGlyAlaProSerGly 740  
Db 3371 TTGGTACGACGACGCGCGAC-----CCGTCCGGA 3342

## RESULT 12

ACA40315

ID ACA40315 standard; DNA; 9903 BP.

XX ACA40315;

XX ACA40315;

DT 19-JUN-2003 (first entry)

XX Prokaryotic essential gene #21972.

XX Antisense; ds; prokaryotic essential gene; cell proliferation;

XX drug design; gene.

XX Mycobacterium tuberculosis.

XX WO20027183-A2.

XX 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 XX  
 XX (ELIT-) ELITRA PHARM INC.  
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX  
 DR WPI: 2003-029926/02.  
 DR P-PSDB; ABU36445.  
 XX  
 PT New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX  
 PS Claim 14; SEQ ID NO 28185; 1766pp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *X. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
 CC prokaryotic essential genes. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 9903 BP; 2129 A; 3380 C; 2727 G; 1667 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 0.018 Length: 9903  
 Score: 192.00 Matches: 194  
 Percent Similarity: 35.04% Conservative: 94  
 Best Local Similarity: 23.60% Mismatches: 323  
 Query Match: 4.76% Indels: 213  
 DB: 7 Gaps: 43

US-09-917-376-3 (1-740) x ACA0315 (1-9903)

Qy 10 AsnValAlaIleGlyGlyGlyPheValAspGlyLeValPheAsnGluGly----- 27  
 Db |||||  
 7399 AACGCGAATCGCGCGCGCAACATCGCGAC-----TTTAACTCGGATCGCA 7449  
 Qy 28 -----AlaProGlyLeuTyrValArgThrAspIleGly-----GlyMet 41  
 Db |||||  
 7450 AACACCGGTCGCGGGCTAACGCGGCTGTCTCAACAAATCGGTATCGGCAACACCGGCAAC 7509

Qy 42 TyrArgTrpAspAlaAlaAsn---GlyArgTrpIleProLeuLeuAspTrpValGlyTrp 60  
 Db |||||  
 7510 TACAACATCGGTTCGGCAACACCGGTAACTACAAC-----ATCGGCTTC 7554  
 Qy 61 AsnAsnTrpGlyTyrAsnGlyVal---ValSerIleAlaAlaAsp----- 74  
 Db |||||  
 7555 GGCAACACCGCGCAACACACATCGGCATCGGCTGTCCGGCGCAACACAGATCGGGTTC 7614  
 Qy 75 ---ProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTyrTrpAsnSerTrpAsp 93  
 Db |||||  
 7615 GGCCCGGTGAACGCGCGC-----ATCGCAACATCGGCTGTTC-----AACCTGGCGGAC 7665  
 Qy 94 ProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrPro 113  
 Db |||||  
 7666 AACAACTTTGCG----- 7677  
 Qy 114 LeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaVal 133  
 Db |||||  
 7678 -----ATGGCCAAACGCGGCAACTTCAACACGAGGCAATGCGC 7713  
 Qy 134 AspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLys---GlyLeuTrp 152  
 Db |||||  
 7714 AACACCGGCAACAAACATCGGTGTTCACACCGCGCAACACACAGTCGGCATCTGG 7773  
 Qy 153 ArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThr 172  
 Db |||||  
 7774 CTGACCGCGGACGGCTTCTCGGCTTCACTCCCTGAACCTCGCGCGCGCAACACCGGT 7833  
 Qy 173 TyrIleAlaAsnProThrAspThrGlyTyrGlnSerAspIleGlnGlyValValTrp 192  
 Db |||||  
 7834 TTCTTCAACTCGGCGACCGCGCAACACCGCGC----- 7863  
 Qy 193 ValAlaPheAspLysSerSerSerLeuGlyGln---AlaSerLysThrIlePheVal 211  
 Db |||||  
 7864 ---TTGTTCACCTCGGCGACCGCGCAACACCGCTGTTCACCTCGGCGACCGCAACGTC 7920  
 Qy 212 GlyValAlaAspProAsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGln 231  
 Db |||||  
 7921 GGATCGCAACATCGGCGACCGCGCGGTTCGGCGCTCGGCGCTATCGGCGGACCGAGTG 7980  
 Qy 232 AlaValProGlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsn 251  
 Db |||||  
 7981 GGATCGGCGGACCAACTCGGCGAGTTTC---AACATCGGCTGTTCACCTCGGCGAC 8037  
 Qy 252 HisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerGly-As 271  
 Db |||||  
 8038 GGCAATGTGCGATCGGCAACTCGGCGACCGCGCAACGTCGGCATCGGCAACACCGGAC 8097  
 Qy 271 pValTrpLysPheSerValThrSerGlyThr----- 282  
 Db |||||  
 8098 GGCAACACCGGCATCGGAACACGCGGCAACTACAACACCGGCTTCTCAACCGGCGCTG 8157  
 Qy 282 p-----ThrArgIleSerProValProSerThrAspThrAlaAsnAs 296  
 Db |||||  
 8158 GTCAACACCGGCATCGGCAACCGGCGCAACCAACACCGGCTGTTCACATCGGCGAC 8217  
 Qy 296 pTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAl 316  
 Db |||||  
 8218 TTCAACACCGGCATCGGCAACCGGCGCACTACAACACCGGCTCTACAACACCGGTAGC 8277  
 Qy 316 aThrGlnIleSerTrp---TrpProAspThrIleIle---PheArgSerThr----- 331  
 Db |||||  
 8278 TACAACACCGGCATGGCAACACCGCGAGACTACGCGCGCGCTTTCATCACCGGCGAC 8337  
 Qy 332 -----AspGlyGlyAlaThrTrpThrArgIleTrpAspTrpTrpSer--- 345  
 Db |||||  
 8338 ATGAACAACCGCTTGTCTCTGCGCGCGCACCGCGGCGGCTGTCTGCGCGCAACTACAC 8397  
 Qy 346 -----TyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaG 360  
 Db |||||  
 8398 ATCAACATCGAGCGACCTGCGCGGTCTCTCAATGTTCGACATCCCGGTCACACATCCCATC 8457  
 Qy 360 uProTrpLeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGlyTr 380

Db 8458 ACCG---CGCATACCAATGCTCCATCCCGCCATTACGTTCCCCAGA-ATC----- 8508  
QY 380 pMetAspGluAlaMetAlaIleAsp---ProPheAsnSerAspArgMetLeuTyrGlyTh 399  
Db 8509 ----GACGCCAGCGAAGGTCGACATAGGCATCTCAGTGGCACCGTCTTGGCCCGGT 8564  
QY 399 rGly---AlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnI 418  
Db 8565 CGGTCCGATCACCTTCGATGGCGGAGCGCGTCGCGCCCGCTGGACACACCCATCGAAAT 8624  
QY 418 eHisIleAlaProMetValLysGlyLeuGluGluThrAlaValAsnAspLeuSerPr 438  
Db 8625 TGACTTCGCGCCC-----TCGCGCGCGATCAACCTCAACATCGCGCA 8666  
QY 438 oProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAs 458  
Db 8667 GCCCGAGCGCTCCACGTGATCAACATCGTGGCGCGCGCGC----- 8709  
QY 458 pValThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerVa 478  
Db 8710 -----GCCGCGCGGATC-----AGCAT 8726  
QY 478 lAspTyrAlaGluLeuAsnProSerIleValArgAlaGlySerPheAspProSerSe 498  
Db 8727 TCGCATCATCGACTTGGCGCCAGCG-----CCCGGCTTCTCAACGCCACAC 8774  
QY 498 rGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnI 518  
Db 8775 CGGCCCG-----TCGTCGGGCTTCTCAACTGGGGTCTGG 8810  
QY 518 ySerGluProGlyGlyValThrThrGlyGlyThr-----ValAlaAl 532  
Db 8811 CAGCGATCGGCTTCTCAACTCGGCAACACTCGGCGCTTACACTTCGCCACTAG 8870  
QY 532 aserAlaAspGlySerArgPheVal-----TrpAlaProGl 544  
Db 8871 CAGCATGGGAATTCGGGCTTCCAAACTATGGTTCGCTGCAGTCGGGCTGGCG- 8925  
QY 544 yAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSer-----Tr 560  
Db 8926 -----AATTTGGGCAACAGCATCTCGGGCATCTA 8954  
QY 560 pAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLy 580  
Db 8955 CAACACCGGCTGGGAGCACCGGCAATGTC-----TCGGGCTTCTCAACATCGG 9005  
QY 580 sThrPheTyrAla-----LeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyVa 598  
Db 9006 CACCAACCTGGTGGTGGTTCGAGAACCGCCG----- 9039  
QY 598 lThrPheGlnProValAlaAlaGlyLeu-ProSerSerGlyAlaValGlyValMetPhe- 617  
Db 9040 -ACCGAGACGACCTTCAGCGTGGGCTTGGCCAACTCGGTTCT---GGAATCTGGGTAG 9095  
QY 618 -----HisAlaValProGlyLysGluGlyAspLeuTrpLeuA 630  
Db 9096 CGCAACATCGCAACTACAACTGGGCGAGCGCAACATCGGCGTCTTCAACACTGGCGAG 9155  
QY 630 la-----AlaSerSerGlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSerA 647  
Db 9156 CGCCACATCGCGGACTTCAACTGGGCGAGCGCAACATCGGCGGACTTCAACTGGCGAG 9215  
QY 647 lalleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerS 667  
Db 9216 CGCCAACA-----TCGGCAGCTCC-AACATCGGTTGGCAACGTCGGTCCGGG--- 9264  
QY 667 erTyrProAlaValPheValGlyThrIleGly---GlyValThrGlyAlaTyrArgS 686  
Db 9265 -----CTGACGGCGGCGCATCGGCAACATCGGCTTCGGCAACACCGGAAACGGAACA 9316  
QY 686 erAspAspCysGlyThrThrTrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnT 706

Db 9317 TCGGCATCGGCAATACCGGCACCGGCACCAATCGGCTTCGGCAACACCGGAACGAAACA 9376  
QY 706 rpGlyGlnAlaIleThrGlyAsp---HisAlaAsnLeuArgValTyrIle-GlyThrA 725  
Db 9377 TCGGCATCGGCTGACCGGCGGACACCATGACCGGGTTTCGGGCTTGAACATCGGCACCG 9436  
QY 725 snGlyArgGlyIleVal-----TyrGlyAspIleGlyGlyAlaProSerGly 740  
Db 9437 GCAACATCGGCTATTCAACTCGGCACCGGCACCAATCGGCTTCGGCAACTCCGCGC 9492

## RESULT 13

AAI99682\_04/c  
Continuation (5 of 45) of AAI99682 from base 400001 (Mycobacterium tuberculosis strain H  
WP Sequence split into 45 fragments LOCUS AAI99682 Accession AAI99682

WP	Fragment Name	Begin	End
WP	AAI99682_00	1	110000
WP	AAI99682_01	100001	210000
WP	AAI99682_02	200001	310000
WP	AAI99682_03	300001	410000
WP	AAI99682_04	400001	510000
WP	AAI99682_05	500001	610000
WP	AAI99682_06	600001	710000
WP	AAI99682_07	700001	810000
WP	AAI99682_08	800001	910000
WP	AAI99682_09	900001	1010000
WP	AAI99682_10	1000001	1110000
WP	AAI99682_11	1100001	1210000
WP	AAI99682_12	1200001	1310000
WP	AAI99682_13	1300001	1410000
WP	AAI99682_14	1400001	1510000
WP	AAI99682_15	1500001	1610000
WP	AAI99682_16	1600001	1710000
WP	AAI99682_17	1700001	1810000
WP	AAI99682_18	1800001	1910000
WP	AAI99682_19	1900001	2010000
WP	AAI99682_20	2000001	2110000
WP	AAI99682_21	2100001	2210000
WP	AAI99682_22	2200001	2310000
WP	AAI99682_23	2300001	2410000
WP	AAI99682_24	2400001	2510000
WP	AAI99682_25	2500001	2610000
WP	AAI99682_26	2600001	2710000
WP	AAI99682_27	2700001	2810000
WP	AAI99682_28	2800001	2910000
WP	AAI99682_29	2900001	3010000
WP	AAI99682_30	3000001	3110000
WP	AAI99682_31	3100001	3210000
WP	AAI99682_32	3200001	3310000
WP	AAI99682_33	3300001	3410000
WP	AAI99682_34	3400001	3510000
WP	AAI99682_35	3500001	3610000
WP	AAI99682_36	3600001	3710000
WP	AAI99682_37	3700001	3810000
WP	AAI99682_38	3800001	3910000
WP	AAI99682_39	3900001	4010000
WP	AAI99682_40	4000001	4110000
WP	AAI99682_41	4100001	4210000
WP	AAI99682_42	4200001	4310000
WP	AAI99682_43	4300001	4410000
WP	AAI99682_44	4400001	4411529

## Alignment Scores:

Pred. No.:	0.285	Length:	110000
Score:	192.00	Matches:	194
Percent Similarity:	35.04%	Conservative:	94
Best Local Similarity:	23.60%	Mismatches:	323
Query Match:	4.76%	Indels:	213
DB:	4	Gaps:	43

US-09-917-376-3 (1-740) x AAI99682\_04 (1-110000)

QY 10 AsnValAlaIleGlyGlyGlyGlyPheValAspGlyIleValPheAsnGlyGly----- 27  
|||||

Db 27279 AACCGCAACATCGCGCGCAACATCGCGAC-----TTTAACGTGGGATCGCA 27229  
Qy 28 -----AlaProGlyIleLeuTyrValArgThrAspIleGly-----GlyMet 41  
Db 27228 AACACCGGTCCGGGGTGTCAACGGCGGTGTCAACACATCGGTATCGGCAACACCGGCAAC 27169  
Qy 42 TyrArgTrpAspAlaAAsn---GlyArgTrpIleProLeuLeuAspTrpValGlyTrp 60  
Db 27168 TACAACATCGGTGTGGCAACACCGGTAACTAACAC-----ATCGGGCTTC 27124  
Qy 61 AsnAsnTrpGlyTyrAsnGlyVal---ValSerIleAlaAlaAsp----- 74  
Db 27123 GGCAACACCGGCAACACACATCGGTATCGGTGTCCGGGCAACACAGATCGGGTTC 27064  
Qy 75 ---ProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAsp 93  
Db 27063 GGCGCGGTGAACCGCGC-----ATCGCAACATGGCGCTGTC-----AACCTGGGGAC 27013  
Qy 94 ProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrPro 113  
Db 27012 AACAACTTTGGC----- 27001  
Qy 114 LeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaVal 133  
Db 27000 -----ATGGCAACCGCGCAACTTCAACAGGGCATTTGCC 26965  
Qy 134 AspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLys---GlyLeuTrp 152  
Db 26964 AACACCGGCAACAAACATCGGTGTTCACACCGGCAACAAACAGTCGGCATCTGG 26905  
Qy 153 ArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThr 172  
Db 26904 CTGACCGGCGACCGGTGTCCGGCTTCAGCTCCCTGAATCTCCGGCGCGCAACACCGGT 26845  
Qy 173 TyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrp 192  
Db 26844 TTCCTCAACTCCGGCAGCGCAACACCGC----- 26815  
Qy 193 ValAlaPheAspLysSerSerSerSerLeuGlyGln---AlaSerLysThrIlePheVal 211  
Db 26814 ---TTGTTCACCTCCGGCACCGGCAACACCGGTGTTCACCTCGGCGCACCGCAACGTC 26758  
Qy 212 GlyValAlaAspProAsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGln 231  
Db 26757 GGCAATCGGCAACATGGCGACCGCGGTTCGGGTTCGGCTATCCGGCGACAGCAGGTG 26698  
Qy 232 AlaValProGlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsn 251  
Db 26697 GGCATCGGCGCACCAACTCGGCGAGTTTC---AACATCGGCTGTTTAACTCGGCGACC 26641  
Qy 252 HisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGly-As 271  
Db 26640 GGCAATGTGGCATCGGCAACTCGGCGACCGGCAACGTCGGCATCGGCAACACCGGCAAC 26581  
Qy 271 pValTrpLysPheSerValThrSerGlyThr----- 282  
Db 26580 GGCAACACCGGCATCGGAACAGCGGCAACTACAAACACCGGTTCCTCAACCGCGGCGCTG 26521  
Qy 282 p-----ThrArgIleSerProValProSerThrAspThrAlaAsnAs 296  
Db 26520 GTCAACACCGGCATCGCAACCGGCGCAACCAACACCGGCGCTGTTCACATCGGCGACC 26461  
Qy 296 pTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAl 316  
Db 26460 TTCAACACCGGCATCGCAACCGGCGCACTACAAACCGGTTCCTTAAACACCGGTAGC 26401  
Qy 316 aThrGlnIleSerTrp---TrpProAspThrIleIle---PheArgSerThr----- 331  
Db 26400 TACAACACCGGCATCGGAACCGGAGACTACGGCACCGCGGTTCATCACCGGCGAGC 26341  
Qy 332 -----AspGlyGlyAlaThrTrpThrArgIleTrpAspTrpThrSer-- 345  
Db 26340 ATGAACAAACGGCTGTCTCGCGCGCGGCGGCGGCGGCTGTGGCGGCAACTACACC 26281

Qy 346 -----TyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGl 360  
Db 26280 ATCAACATCGAGCAGCTCGCGGTCTCTCAATGTGACATCCCGTCAACATCCCATC 26221  
Qy 360 uproTrpLeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGlyTr 380  
Db 26220 ACCG---GCGACATCAACCAATGTCTCCATCCCGCATTTACGTTCGCCAGA-ATC----- 26170  
Qy 380 pMetAspGluAlaMetAlaIleAsp---ProPheAsnSerAspArgMetLeuTyrGlyTh 399  
Db 26169 ---GAGCCAGCGGAGCGTCGACATAGGCATCTCAGTCGACCGGTTCGCCCGGT 26114  
Qy 399 rGly---AlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlnI 418  
Db 26113 CGGTCCGATCACCTCGATGCGGGGACGCGCGCGCTGGACACACCCATCGAAAT 26054  
Qy 418 eHisIleAlaProMetValLysGlyLeuGluGluThrAlaValAsnAspLeuIleSerPr 438  
Db 26053 TGACTTCGGCCCC-----TCGCGCGGATCAACCTCAACATCGGCAA 26012  
Qy 438 oProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAs 458  
Db 26011 GCCGACGGCTCCACCGGTGATCAACATCGTGGCGCGCGCGC----- 25969  
Qy 458 pValThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerVa 478  
Db 25968 -----GCCGCGCGATC-----AGCAT 25952  
Qy 478 lAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerSe 498  
Db 25951 TCCGATCATCGACTTCGGCGCAGC-----CCCGCTCTTCTCAACGCCACCA 25904  
Qy 498 rGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnG 518  
Db 25903 CGGCGCG-----TCGTCGGCTCTCTCACTCGGCTGCTGG 25868  
Qy 518 ySerGluProGlyGlyValThrThrGlyGlyThr-----ValAlaAl 532  
Db 25867 CAGCGCATCGGCTGTCTGAATTCGCAACAACTCGGCGCTTACAACTTCGCCACTAG 25808  
Qy 532 aSerAlaAspGlySerArgPheVal-----TrpAlaProGl 544  
Db 25807 CAGCATGGGAATTCGGGCTTCCAAAATATGGGTCTGTCAGTCGGGTGGCGC----- 25753  
Qy 544 yAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSer-----Tr 560  
Db 25752 -----AATTGGGCAACAGCATCTCGGCGCATCTA 25724  
Qy 560 pAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLy 580  
Db 25723 CAACACCGGTTCGGAGCAGCGGCAATGTC-----TCGGGTCTCTCAACATCGG 25673  
Qy 580 sThrPheTyrAla-----LeuSerAsnGlyThrPheTyrArgSerThrAspGlyVa 598  
Db 25672 CACCAACCTGGCTGGGTGGTTCGACACCGCGC----- 25639  
Qy 598 lThrPheGlnProValAlaAlaGlyLeu- ProSerSerGlyAlaValGlyValMetPhe- 617  
Db 25638 -ACGAGACGACCTTCAGCGGTGGGCTTCGCAACCTCGGCTTCT---GGAAATCTGGGTAG 25583  
Qy 618 -----HisAlaValProGlyLysGluGlyAspLeuTrpLeuA 630  
Db 25582 CGCAACATCGGCAACTACAACTGGGCGAGCGGCAACATCGGCGCTCTACAACTGGGCGAG 25523  
Qy 630 la-----AlaSerSerGlyLeuTyrHisSerThrAsnGlyGlySerTrpSera 647  
Db 25522 CGCAACATCGGCGACTTCAACCTGGGCGAGCGGCAACATCGGCGACTTCAACCTGGGCGAG 25463  
Qy 647 lalleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerS 667  
Db 25462 CGCAACA-----TCGCGAGCTCC-AAACATCGGTTTCGGCAACGTCGTCGCGG----- 25414



QY 667 erTyrProAlaValPheValGlyThrIleGly---GlyValThrGlyAlaTyrArgS 686  
 Db 25413 -----CTGACGGCGGCATCGGCAACATCGGCTTCGGCAACACCGGAAACGGAACA 25362  
 QY 686 erAspAspCysGlyThrTrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnT 706  
 Db 25361 TCGGCATCGGCATATCCGGCACCGCAACATCGGCTTCGGCAACACCGGAAACGGAACA 25302  
 QY 706 rpGlyGlnAlaIleThrGlyAsp--HisAlaAsnLeuArgValTyrIle-GlyThrA 725  
 Db 25301 TCGGCATCGGCTGACCGCGCACCATGACCGGTTTCGGCGCTGGAACATCGGCACCG 25242  
 QY 725 snGlyArgGlyIleVal-----TyrGlyAspIleGlyAlaProSerGly 740  
 Db 25241 GCAACATCGGCTATTCAACTCGCGCACCGGCAACATCGGCTTCGGCAACATCGGCGC 25186

## RESULT 14

ACA38418  
 ID ACA38418 standard; DNA; 3324 BP.

XX ACA38418;  
 XX  
 DT 19-JUN-2003 (first entry)  
 XX  
 DE Prokaryotic essential gene #20075.  
 XX  
 KW Antisense; ds; prokaryotic essential gene; cell proliferation;  
 KW drug design; gene.  
 XX

OS Mycobacterium bovis.  
 XX  
 PN WO200277183-A2.  
 XX  
 PD 03-OCT-2002.  
 XX

PF 21-MAR-2002; 2002WO-US009107.  
 XX  
 PR 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen XL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX  
 DR WPI; 2003-029926/02.  
 DR P-PSDB; ABU34548.  
 XX

PT New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX  
 PS Claim 14; SEQ ID NO 26288; 1766pp; English.  
 XX

CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an

CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
 CC prokaryotic essential genes. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

SQ Sequence 3324 BP; 726 A; 1184 C; 886 G; 528 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 0.0169 Length: 3324  
 Score: 184.00 Matches: 205  
 Percent Similarity: 35.59% Conservative: 89  
 Best Local Similarity: 24.82% Mismatches: 313  
 Query Match: 4.56% Indels: 221  
 DB: 7 Gaps: 48

US-09-917-376-3 (1-740) x ACA38418 (1-3324)

QY 10 AsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsnGlyGly----- 27  
 Db 181 AACCGCAACATCGCGCGCGCAACATCGGCAC-----TTTAACGTCGGGATCGCA 231  
 QY 28 -----AlaProGlyIleLeuTyrValArgThrAspIleGly-----GlyMet 41  
 Db 232 AACACCGCTCGGCGCTAAACGCGCGTGTCAACATCGGTATCGGCAACACCGGCAAC 291  
 QY 42 TyrArgTrpAspAlaAlaAsn--GlyArgTrpIleProLeuLeuAspTrpValGlyTrp 60  
 Db 292 TACAACATCGGTGCGGCACACACCGGTAACTACAAC-----ATCGGCTTC 336  
 QY 61 AsnAsnTrpGlyTyrAsnGlyVal---ValSerIleAlaAlaAsp----- 74  
 Db 337 GGCAACACCGCGCAACACATCGGCATCGCGCTGTCCGCGCACACACAGATCGGGTTC 396  
 QY 75 ---ProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAsp 93  
 Db 397 GGCCCGCTGAACGCGCGC-----ATCGCAACATCGGCTGTTC---AACCTGGGCGAC 447  
 QY 94 ProAsnAspGly---AlaIleLeuArgSerSerAspGlnGlyAla-----ThrTrpGln 110  
 Db 448 AACACATTTGGCATGGCCAACCGCGGCAACTTCAACAGGCGCATTCACACCGGCAAC 507  
 QY 111 IleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArg 130  
 Db 508 AACCAACATCGGCTGTTCACACCGCGCAACACACATCGCGCATCGGCTG----- 558  
 QY 131 LeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyGly 150  
 Db 559 -----ACCGGCAACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 570  
 QY 151 Leu-----TrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPhePro 168  
 Db 571 TTGTCCGGCTTCACTCCCTGAACTCCGGCGCC-----GGCAACACCGGTTTCTTC 621  
 QY 169 AspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGln 188  
 Db 622 AACTCCGGCAC-----GCCAAC-----ACCGGCG----- 645  
 QY 189 GlyValValTrpValAlaPheAspLysSerSerSerSerLeuGlyGln---AlaSerLys 207  
 Db 646 -----TTGTTCACTCGCGCGCACCGCGCAACACCGGCTTGTTCACATCGGCGC 690  
 QY 208 ThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSerArgAspGlyGly 227

Db 691 ACCGGCAACGTCGCGATCGGCAACATCGGCAACCGCGGCTTCGGCGTCGGCCTATCCGGC 750  
 Qy 228 AlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIleProHisLysGlyValPhe 247  
 Db 751 GACAGCCAGGTGGCGATCGGCGGACCACTCGGCGAGTTTC---AACATCGGCTTGT 807  
 Qy 248 AspProValAsnHisValLeuTyrlleAlaThrSerAsnThrGlyGlyProTyAspGly 267  
 Db 808 AACTCGGCGCACCGGCAATGTCGGCATCGGCAACTCGGCGACCGCAACGTCGGCATCGGC 867  
 Qy 268 SerSerGlyAspValTrpLysPheSerValThrSerGlyThr----- 281  
 Db 868 AACACCGGCGACCGGCAACCGGATCGGAACAGCGGCAACTACACACCGGCTTGCTC 927  
 Qy 282 -----Trp-----ThrArgIleSerProValProSerThrAs 292  
 Db 928 AACCGGCGCTGTGTCAACACCGGATCGCCNACCGGCGCAACCAACACCGGCTGTTC 987  
 Qy 292 pThrAlaAsnAspTyrlleGlyTyrlleGlyLeuThrIleAspArgGlnHisProAsnTh 312  
 Db 988 AACATCGGCGACCTTCAACACCGGATCGCCNACCGGCGCAACTACACACCGGCTCTAC 1047  
 Qy 312 rIleMetValAlaThrGlnIleSerTrp---TrpProAspThrIleIle---PheArgSe 330  
 Db 1048 AACACCGGCTAGTACACACCGGATCGGCAACCGGAGACTACGCGCGCGCTTC 1107  
 Qy 330 rThr-----AspGlyGlyAlaThrTrpThrArgIleTrpAs 342  
 Db 1108 ATACCGGCGCATGAACACCGGCTTCTCTGGCGCGCGGACCGGCGGCTCTGGCG 1167  
 Qy 342 pTrpThrSer-----TyrlleProAsnArgSerLeuArgTyrlleValLeuAs 356  
 Db 1168 GCCAACTACACCATCACCATCGAGCGACTCGCGCTTCTCAATGTCGACATCCCGGTC 1227  
 Qy 356 pIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProProValProSerPr 376  
 Db 1228 AACATCCCGCATCCCG---CGGACATCACCATCTCTCCATCCCGCGCATTTACCTCC 1284  
 Qy 376 oLysLeuGlyTrpMetAspGluAlaMetAlaIleAsp---ProPheAsnSerAspArgMe 395  
 Db 1285 AGA-ATC-----GACCGCGGAGGAGTGCATAGGATCTCTCATGTCGACCGT 1334  
 Qy 395 tLeuTyrlleGlyThrGly---AlaThrLeuTyrlleAlaThrAsnAspLeuThrIleTyrlle 414  
 Db 1335 CTGGCGCGCGTCCGTCGATCACCTGTCATGGCGGCGGACGCTCGCGCGCGTGACAC 1394  
 Qy 414 rGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluThrAlaValAsnAs 434  
 Db 1395 ACCCATCGAAATTGACTTCGGCGCC-----TCGCGCGCGATCAACCT 1436  
 Qy 434 pLeuIleSerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPh 454  
 Db 1437 CAACATCGGCAAGCCGCGCTCCACCGTGCATCAACATCGGCGCGCGCGCGC----- 1491  
 Qy 454 eThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrTh 474  
 Db 1492 -----GCGCGCGCGATC----- 1503  
 Qy 474 rGlyThrSerValAspTyrlleAlaGluLeuAsnProSerIleIleValArgAlaGlySerPh 494  
 Db 1504 -----AGCATTCGATCATCGACTTCGGCGCACG-----CCGCGCTTCTT 1544  
 Qy 494 eAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAs 514  
 Db 1545 CAACCGCACCGCGCGCG-----TCGCGCGCTTCTTCAA 1580  
 Qy 514 nTrpPheGlnGlySerGluProGlyGlyValThrThrGlyGlyThr----- 529  
 Db 1581 CTGGGTGCTGGAGCGCATCGGCTTCTGAACTTCGGCAACAACTCGGCGCTCTACAA 1640  
 Qy 530 ----ValAlaAsnAlaAspGlySerArgPheVal----- 540

Db 1641 CTTGCCCACTAGCAGCATGGGAATTCGGGCTTCAAAACCTATGGGTCTCGCTAGTCGGG 1700  
 Qy 541 -TrpAlaProGlyAspProGlyGlnProValValTyrlleAlaValGlyPheGlyAsnSer-- 559  
 Db 1701 CTGGGCG-----AATTGGGCAACAGCAT 1724  
 Qy 560 -----TrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspAr 576  
 Db 1725 CTCGGGATCTACAAACACCGCTTGGGAGCACCGGCAAAATGTC-----TCGGGCTT 1775  
 Qy 576 gValAsnProLysThrPheTyrlleAla-----LeuSerAsnGlyThrPheTyrlleArgSerTh 594  
 Db 1776 GCTCAACATCGCGCAACCTGGGTGGGTGTGCAGAACCGGCGC----- 1821  
 Qy 594 rAspGlyGlyValThrPheGlnProValAlaAlaGlyLeu-ProSerSerGlyAlaValG 614  
 Db 1822 -----ACCGAGACGACTTCAGGTGGCTTGGCAACCTCGGCTCT- -G 1865  
 Qy 614 lValMetPhe-----HisAlaValProGlyLysGluGlyA 626  
 Db 1866 GAATCTGGGTAGCGCAACATCGGCACTACAACTGGGCGCGGCAACATCGGCTCTA 1925  
 Qy 626 spLeuTrpLeuAla-----AlaSerSerGlyLeuTyrlleHisSerThrAsnGlyGlyS 643  
 Db 1926 CAACCTGGGCGAGCGCAACATCGGCGACTTCAACCTGGGCGAGCGCAACATCGGCGACTT 1985  
 Qy 643 erSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerA 663  
 Db 1986 CAACCTGGGCGAGCGCAACA-----TCGGCAGCTCC-AACATCGGTTTCGGCAACGCTCG 2038  
 Qy 663 laProGlySerSerTyrlleProAlaValPheValValGlyThrIleGly---GlyValThrG 682  
 Db 2039 GTCCGGG-----CTGACGGCGGCGCATCGGCAACATCGCTTCGGCGCAACACCG 2086  
 Qy 682 lValAlaTyrlleArgSerAspAspCysGlyThrThrTrpValLeuIleAsnAspAspGlnHisG 702  
 Db 2087 GAAACGGAACATCGGCAATACCGGCAATACCGGCAACATCGGCTTCGGCAACACCG 2146  
 Qy 702 lntTyrlleAsnTrpGlyGlnAlaIleThrGlyAsp---HisAlaAsnLeuArgArgValTy 721  
 Db 2147 GAAACGGAACATCGGCAATCGGCTCGGCTGACCGGCGACCACTGACCGGTTTCGGCGCTGA 2206  
 Qy 721 rIle-GlyThrAsnGlyArgGlyIleVal-----TyrlleAspIleGlyGlyAlaP 738  
 Db 2207 ACTCGGCGACCGCAACATCGGCTTATCAACTCGGCGACCGCAACATCGGCTTCGGCA 2266  
 Qy 738 roSerGly 740  
 Db 2267 ACTCGGCG 2274  
 RESULT 15  
 AAI99683\_04/c  
 Continuation (5 of 44) of AAI99683 from base 400001 (Mycobacterium tuberculosis strain H  
 WP Sequence split into 44 fragments LOCUS AAI99683 Accession AAI99683  
 WP Fragment Name Begin End  
 WP AAI99683\_00 1 110000  
 WP AAI99683\_01 100001 210000  
 WP AAI99683\_02 200001 310000  
 WP AAI99683\_03 300001 410000  
 WP AAI99683\_04 400001 510000  
 WP AAI99683\_05 500001 610000  
 WP AAI99683\_06 600001 710000  
 WP AAI99683\_07 700001 810000  
 WP AAI99683\_08 800001 910000  
 WP AAI99683\_09 900001 1010000  
 WP AAI99683\_10 1000001 1110000  
 WP AAI99683\_11 1100001 1210000  
 WP AAI99683\_12 1200001 1310000  
 WP AAI99683\_13 1300001 1410000  
 WP AAI99683\_14 1400001 1510000  
 WP AAI99683\_15 1500001 1610000  
 WP AAI99683\_16 1600001 1710000  
 WP AAI99683\_17 1700001 1810000



```
Db 26073 CAAGCCACACCGCGCGG-----TCGTCGGGCTTCCTCAA 26038
Qy 514 nTrpPheGlnGlySerGluProGlyValThrThrGlyGlyThr----- 529
Db 26037 CTGGGGTGCTGGCAGCGCATCGGGCTTGCTGAACCTTCGGCAACAACCTCGGGCCTCTACAA 25978
Qy 530 -----ValAlaAlaSerAlaAspGlySerArgPheVal----- 540
Db 25977 CTTGGCCACTAGCAGCATGGGAAATTCGGGCTTCAAAACTATGGGTGCTGCAGTCGGG 25918
Qy 541 -TrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSer-- 559
Db 25917 CTGGGCG-----AATTGGGCAACAGCAT 25894
Qy 560 -----TrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspAr 576
Db 25893 CTGGGGCATCTACACACCGGCTTGGGAGCAGCAGCGCAATGTC-----TCGGGCTT 25843
Qy 576 gValAsnProGlyThrPheTyrAla-----LeuSerAsnGlyThrPheTyrArgSerTh 594
Db 25842 GCTCAACATCGGCACCAACCTGGTGGTGGTTCGACAGCGGCG----- 25797
Qy 594 xAspGlyGlyValThrPheGlnProValAlaAlaGlyLeu-ProSerSerGlyAlaValG 614
Db 25796 -----ACGAGAGACCTTCAGCGTGGGCTTGGCCAACTCGGGTTCT--G 25753
Qy 614 lyValMetPhe-----HisAlaValProGlyLyGluGlyA 626
Db 25752 GAATCTGGGTAGCGCAACATCGGCAACTACACCTGGGCGCGCCMACATCGGCGTCTA 25693
Qy 626 sPLeuTrpLeuAla-----AlaSerSerGlyLeuTyrHisSerThrAsnGlyGlys 643
Db 25692 CAACCTGGGCGAGCGCAACATCGGCGACTTCAACCTGGGCGAGCGCAACATCGGCGACTT 25633
Qy 643 erSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerA 663
Db 25632 CAACCTGGGCGAGCGCAACA-----TCGGCAGCTCC-AAACATCGGGTTCGGCAACGTCG 25580
Qy 663 laProGlySerSerTyrProAlaValPheValValGlyThrIleGly---GlyValThrG 682
Db 25579 GTCGGGG-----CTACGCGCGCCATCGGCAACATCGGCTTCGGCAACACCG 25532
Qy 682 lyAlaTyrArgSerAspAspCysGlyThrThrTrpValLeuIleAsnAspAspGlnHisG 702
Db 25531 GAAACGAAACATCGGCATCGGCAATACCGGCACCGCAACATCGGCTTCGGCAACACCG 25472
Qy 702 lnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAsp--HisAlaAsnLeuArgArgValTy 721
Db 25471 GAAACGAAACATCGGCATCGGCTCGGCTGACCGCGCACACCATGACCGGGTTCGGCGGTGGA 25412
Qy 721 rIle-GlyThrAsnGlyArgGlyIleVal-----TyrGlyAspIleGlyGlyAlap 738
Db 25411 ACTCGGGCACCGCAACATCGGGCTATTCAACTCGGCGACCGGCAACATCGGCTTCGGCA 25352
Qy 738 roSerGly 740
Db 25351 ACTCGGCG 25344
```

Search completed: May 12, 2004, 05:39:13  
Job time : 1279.48 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 12, 2004, 04:33:52 ; Search time 6950.51 Seconds  
(without alignments)  
3179.336 Million cell updates/sec

Title: US-09-917-376-3  
Perfect score: 4036  
Sequence: 1 ATTQPTWNSNAIVGGGFGVD.....YIGTNGRGIYVGDIGGAPSG 740

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_epool/US0917376/runat\_11052004\_114535\_28249/app\_query.fasta\_1.1429  
-DB=EST -QFMT=fastcap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US0917376@cgn\_1\_1\_5334 @runat\_11052004\_114535\_28249 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPOP=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_man:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rdg:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vri:\*  
28: gb\_gssI:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	615.5	15.3	704	14	CF872104	CF872104 trico29xo
2	615.5	15.3	761	14	CB902332	CB902332 trico29xo
3	587.5	14.6	747	14	CF876916	CF876916 trico74xd
4	587.5	14.6	814	14	CB905388	CB905388 trico74xd
5	415.5	10.3	707	14	CF880713	CF880713 trico82xn
6	415.5	10.3	782	14	CB907625	CB907625 trico82xn
7	385.5	9.6	693	14	CF882065	CF882065 trico29xo
8	361.5	9.0	751	14	CF868882	CF868882 trico16xm
9	361.5	9.0	803	14	CB898982	CB898982 trico16xm
10	338	8.4	738	14	CF866388	CF866388 trico06xe
11	338	8.4	794	14	CB896354	CB896354 trico06xe
12	314.5	7.8	546	10	BF072664	BF072664 NCMS3H773
13	306.5	7.6	389	13	BU639045	BU639045 mgcw011xd
14	286.5	7.1	713	14	CF875815	CF875815 trico39xx
15	286.5	7.1	782	14	CB904767	CB904767 trico39xx
16	282.5	7.0	719	14	CF881775	CF881775 trico85xf
17	282.5	7.0	775	14	CB908435	CB908435 trico85xf
18	278.5	6.9	929	14	CF885920	CF885920 trico85xf
19	222	5.5	796	14	CF391910	CF391910 RTDR3_10
20	183	4.5	3057	11	AK044947	AK044947 Mus muscu
21	180	4.5	4355	29	AY416870	AY416870 Mus muscu
22	171.5	4.2	296	13	BU641770	BU641770 mgmk008xa
23	170	4.2	4631	11	AK082944	AK082944 Mus muscu
24	166.5	4.1	338	9	AW064456	AW064456 SP1055 KR
25	153.5	3.8	3623	11	AK083848	AK083848 Mus muscu
26	153.5	3.8	3751	11	AK080277	AK080277 Mus muscu
c 27	153.5	3.8	4101	29	AY402312	AY402312 Homo sapi
28	151	3.7	4654	11	AK086725	AK086725 Mus muscu
c 29	148.5	3.7	1178	28	CC203127	CC203127 CH261-8B4
30	140	3.5	3782	11	AK041860	AK041860 Mus muscu
c 31	139.5	3.5	5367	11	AK084803	AK084803 Mus muscu
c 32	138	3.4	5073	29	AY407742	AY407742 Homo sapi
33	137.5	3.4	1886	11	AK009469	AK009469 Mus muscu
34	137	3.4	1581	11	BC029959	BC029959 Homo sapi
35	136.5	3.4	1569	29	AY407967	AY407967 Homo sapi
36	136	3.4	1674	29	AY412937	AY412937 Mus muscu
37	134	3.3	426	13	BQ751758	BQ751758 EST632321
38	133.5	3.3	2545	14	CD014011	CD014011 90138962
39	132	3.3	2565	11	AK030673	AK030673 Mus muscu
40	132	3.3	3264	29	AY404484	AY404484 Homo sapi
41	132	3.3	4611	29	AY418194	AY418194 Homo sapi
42	131.5	3.3	3762	11	BC039264	BC039264 Homo sapi
43	131.5	3.3	4132	11	AK032858	AK032858 Mus muscu
c 44	131.5	3.3	6914	11	BC042295	BC042295 Homo sapi
45	130	3.2	1723	11	AK089973	AK089973 Mus muscu

ALIGNMENTS

RESULT 1  
CF872104  
LOCUS CF872104 704 bp mRNA linear EST 31-OCT-2003  
DEFINITION trico29xo13.b1 T.reesei mycelial culture, Version 6 October 2003  
ACCESSION CF872104  
VERSION CF872104.1 GI:38126786  
KEYWORDS EST.  
SOURCE Hypocrea jecorina (anamorph: Trichoderma reesei)  
ORGANISM Hypocrea jecorina  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.  
REFERENCE 1 (bases 1 to 704)

**AUTHORS** Diener, S.E., Dankmeyer, L., Dunn-Coleman, N., Houfek, T.D., Mitchell, T.K., van Solingen, P., Teunissen, P.J.M., Ward, M. and Dean, R.A.  
**TITLE** Analysis of the protein processing and secretion pathways in a Trichoderma reesei EST dataset  
**JOURNAL** Unpublished (2003)  
**COMMENT** Contact: Ralph A. Dean  
Fungal Genomics Laboratory  
North Carolina State University  
Campus Box 7251, Raleigh, NC 27695, USA  
Tel: 919-513-0020  
Fax: 919-513-0024  
Email: ralph\_dean@ncsu.edu  
Seq primer: LT-F1 primer.  
**FEATURES**  
source  
1..704  
/organism="Hypocrea jecorina"  
/mol\_type="mRNA"  
/strain="QM6a"  
/db\_xref="taxon:51453"  
/clone="tric029x013"  
/dev\_stage="mycelia"  
/clone\_lib="T.reesei mycelial culture, Version 6 October 2003"  
/notes="Vector: pREP3Y; Site\_1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."  
**ORIGIN**  
Alignment Scores:  
Pred. No.: 2,4e-43 Length: 704  
Score: 615.50 Matches: 118  
Percent Similarity: 75.61% Conservative: 37  
Best Local Similarity: 57.56% Mismatches: 43  
Query Match: 15.25% Indels: 9  
DB: 14 Gaps: 4  
US-09-917-376-3 (1-740) x CF872104 (1-704)  
QY 6 TyrThrTrpSerAsnValAlaIle---GlyGlyGlyGlyPheValAspGlyIleValPhe 24  
DB 98 TTTTTCATGAAGAAGCTCAGCTCGCGCGCGCGCTTCGTCGCCGATCATCTTC 157  
QY 25 AsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArg-- 43  
DB 158 CATCCCAAGACAAAGGCTAGCATATGCACGAACAGATATGGGGCTGTACCGCTC 217  
QY 44 TrpAspAlaAlaAsnGlyArgTrpIlePro-----LeuLeuAspTrpValGlyTrpAsn 61  
DB 218 TTTCCTCGGACTCATGGACCGCC-GTCACGGATGGGATTGCTGATAATGCCGGCTGGCAC 276  
QY 62 AsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysVal 81  
DB 277 AACTGG-----GGCATCGACCTGTGTGGCTTGCAGCGAGCATCAAAAGGTG 327  
QY 82 TrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArg 101  
DB 328 TATCGCGAGTCGCGATGTATACGAACAGCTGGATCGAGTAATGGAGCATCATTCGC 387  
QY 102 SerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsn 121  
DB 388 TCGTCAGACCGCGGCGCAACGTGTCTTCAACCACTTGCCTTCAAAAGTCGGGGTAAC 447  
QY 122 MetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeu 141  
DB 448 ATGCAGAGACGGAGACCGGAGAGCGTCTGGCTGTGTCATCGCGCAACTCCAAATCATC 507  
QY 142 TyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrp 161  
DB 508 TACTTTGGTCTCGCTCAGGAACCGGCTCTGGAAGTCTACGAGCGCGGCTGACCTTT 567  
QY 162 SerGlnMetThrAsnPheProAspValGlyThrTyrlleAlaAsnProThrAspThrThr 181







```

Db 213 TCTGACCCGCGGAGACGAGCTATTGGCCGACGTCGGAGACGACACGCTTTCACCTTT 272
Qy 457 Ala-----AspValThrAlaValProSerThrPheThrSerProValPheThr 473
Db 273 GCCAGCGAAGACACCTCGGACATCCCGACAGCGTCTGGGCAACGCCACATGGGCC 332
Qy 474 ThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleValAlaGlySer 493
Db 333 ACCTCGACGAGCGTCGACTACGCGCGGAACTCGTCAAGAGCGTCTGTCGCGTCGGCAAC 392
Qy 494 PheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLys 513
Db 393 ACCGCCGCGACCAACAG-----GTGGCCATCTCTGTCGACGCGCGCGCG 437
Qy 514 AsnTrpPheGlnGlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSer 533
Db 438 ACGTGGAGCATCGACTACGCGCGCGACACGTCATGAACGCGCGCACGGTGGCTATTTCG 497
Qy 534 AlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAla 553
Db 498 GCCAGCGCGACACGATCTCTGGTCACCGCTCTGTCGCG-----GTGCAGCGC 548
Qy 554 ValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArg 573
Db 549 TCGCAGTTCCAGGCGAGCTTTCGCTCGCTCGAGCCCTGCGCGCGCGCGCTCATCGCC 608
Qy 574 SerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSer 593
Db 609 TCGGACAAAGAACCAACACGCGTCTTCTACGCGGCTCCGATCGACCTTTTACGTGAGC 668
Qy 594 ThrAspGlyGlyValThrPhe 600
Db 669 AAGGACACCGCGACGACTTC 689

```

```

RESULT 6
CB907625          782 bp  mRNA  linear  EST 02-JUL-2003
LOCUS             trico02xn19 T.reesei mycelial culture, Version 3 april Hypocrea
DEFINITION        jecorina cDNA clone trico02xn19, mRNA sequence.
ACCESSION         CB907625
VERSION           CB907625.1 GI:30122283
KEYWORDS
SOURCE            EST.
ORGANISM          Hypocrea jecorina (anamorph: Trichoderma reesei)

```

```

REFERENCE
AUTHORS           Foreman,P.K., Brown,D.E., Dankmeyer,L., Dean,R., Diener,S.,
                  Dunn-Coleman,N.S., Goedegebuur,F., Houfek,T.D., England,G.J.,
                  Kelley,A.S., Meerman,H.J., Mitchell,T., Mitchinson,C.,
                  Olivares,H.A., Teunissen,P.J., Yao,J. and Ward,M.
TITLE             Transcriptional regulation of biomass-degrading enzymes in the
                  filamentous fungus Trichoderma reesei
JOURNAL           J. Biol. Chem. 278 (34), 31988-31997 (2003)
MEDLINE           22803314
PUBMED            12788920
COMMENT           Contact: Pamela K. Foreman
                  Genencor Intl.
                  925 Page Mill Road, Palo Alto, CA 94304, USA
                  Tel: (650) 846-7635
                  Fax: (650) 621-7817
                  Email: Pforeman@genencor.com
                  Seq primer: LT-F1 primer.
                  Location/Qualifiers

```

```

FEATURES
source            1..782
                  /organism="Hypocrea jecorina"
                  /mol_type="mRNA"
                  /strain="QM6a"
                  /db_xref="taxon:51453"
                  /clone="trico02xn19"
                  /dev_stage="mycelia"
                  /clone_lib="T.reesei mycelial culture, Version 3 april"

```

```

/note="vector: pREP3Y, Site_1: Not I/Sal I; Mycelial
culture grown from 24 hrs to 6 days with varying Carbon
and Nitrogen sources and concentrations."

ORIGIN
Alignment Scores:      8.7e-26      Length:      782
Pred. NO.:            415.50      Matches:      92
Score:                55.51%      Conservative: 34
Percent Similarity:    40.53%      Mismatches:   90
Best Local Similarity: 10.29%      Indels:       11
Query Match:          14          Gaps:         3
DB:

US-09-917-376-3 (1-740) x CB907625 (1-782)

Qy 377 LysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeu 396
Db 108 CGCTTCGGCTGGATTTGAGTCTCTCGAGATTGACCAACCGACAGCAACCACTGGCTC 167
Qy 397 TyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGly 416
Db 168 TAGCGCACCGGATACACAATCTTTGGCGGCCACGATCTCACCACTGGGACACGCGCCAC 227
Qy 417 GlnIleHisIleAlaProMetValLysGlyLeuGluGluThrAlaValAsnAspLeuIle 436
Db 228 AATGTGTAATCCAATCAATCACTGCGACAGCGCATTCGAGGAATTCCTCGTCCAGGACCTGGCC 287
Qy 437 SerProProSerGlyAlaProIleLysSerAlaLeuGlyAspLeuGlyGlyPheThrHis 456
Db 288 TCTGCACCCCGCGAAGCGAGCTATTGGCCGAGTCGGAGACGACCAACGGCTTCACCTTT 347
Qy 457 Ala-----AspValThrAlaValProSerThrIlePheThrSerProValPheThr 473
Db 348 GCCAGCAGAAACGACCTCGGACATCGCGCAGACGGTCTGGGCAACGCCACATGGGCC 407
Qy 474 ThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleValAlaGlySer 493
Db 408 ACCTCGACGAGCGTCGACTACGCGCGGAACCTCGGTCAGAGCGTCTGCGCGTCGGCAAC 467
Qy 494 PheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLys 513
Db 468 ACCGCCGCGACGCAACAG-----GTGGCCATCTCTGTCGACGCGCGCGCG 512
Qy 514 AsnTrpPheGlnGlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSer 533
Db 513 ACGTGGAGCATCGACTACGCGCGCACGCTCATGAACGCGCGCACGCTGGCTATTTCG 572
Qy 534 AlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAla 553
Db 573 GCCAGCGCGACACGATCTCTGGTCGACCGCTCTGTCGCG-----GTGCAGCGC 623
Qy 554 ValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArg 573
Db 624 TCGCAGTTTCCAGGCGAGCTTTGCTCTCGAGCTGTCGAGCGTCCGCGCGCGCGCTCATCGCC 683
Qy 574 SerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSer 593
Db 684 TCGGACAAAGAACGACCAACAGCGCTCTTACGCGCGCTCCGGATCGGACCTTTTACGTGAGC 743
Qy 594 ThrAspGlyGlyValThrPhe 600
Db 744 AAGGACACCGCGACGACTTC 764

RESULT 7
CF882065          693 bp  mRNA  linear  EST 31-OCT-2003
LOCUS             trico29xn03.b11 T.reesei mycelial culture, Version 6 October 2003
DEFINITION        Hypocrea jecorina cDNA clone trico29xn03, mRNA sequence.
ACCESSION         CF882065
VERSION           CF882065.1 GI:38136747
KEYWORDS          EST.
SOURCE            Hypocrea jecorina (anamorph: Trichoderma reesei)
ORGANISM          Hypocrea jecorina

```



504	Qy	HisValAlaPheSerThrAspGlyGlyLeuAsnTrpPheGlnGlySerGluProGlyGly	523
303	Db	---GTGGCCATCTCGTCGACGGCGCGACGTGGAGCATCGCACTACGGCGCGCACACG	359
524	Qy	ValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPro	543
360	Db	TCCATGAACGGCGGCACGGTGGCTATTTCGGCCGACGGCGACAGATCCCTCTGGTCGACC	419
544	Qy	GlyAspProGlyGlnProValValThrAlaValGlyPheGlyAenSerTrpAlaAlaSer	563
420	Db	GCCTCGTCGGGC-----GTGCAGCGCTCGCAGTTCCAGGCGACGTTTGCCCTCCGTC	470
564	Qy	GlnGlyValProAlaAenAlaGlnIleArgSerAspArgValAenProLysThrPheTyr	583
471	Db	TCGAGCTTCGCCCGGGCGCGTCATCGCTCGGACAAAGAACCAACAGCGCTCTTCAC	530
584	Qy	AlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProVal	603
531	Db	GCCGGTCCGGATCGACCTTTTACGTCCAGCAAGGACACCGGCGACAGCTC-----	581
604	Qy	AlaAlaGlyLeuProSerSerGlyAlaValGlyValMet-----PheHisAlaValPro	621
582	Db	ACGCGCGGG---CCCAAGCTGGCGACGGCGAGGAGATCGCGGATATCGCTGCTCACCCG	638
622	Qy	GlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsnGly	641
639	Db	ACCACCGGGCACGTTGTATGTTCTGCACCGACGTCGGCATATTCCGCTCCACAGACTCG	698
642	Qy	GlySerSerTrpSerAlaIleThr	649
699	Db	GGCAGCAGCTTTGGCCAAAGTCTCC	722

RESULT 9				
CB989882				
LOCUS	CB989882	803 bp	mRNA	linear
DEFINITION	trio016xm14 T.reesei mycelial culture, Version 3 april Hypocrea jecorina cDNA clone trio016xm14, mRNA sequence.			

ORIGIN	
Alignment Scores:	
Pred. No.:	4.78e-21
Score:	361.50
Percent Similarity:	51.61%
Best Local Similarity:	35.48%
Query Match:	8.96%
DB:	14
	Length:
	Matches:
	Conservative:
	Mismatches:
	Indels:
	Gaps:
	803
	88
	40
	103
	17
	6
US-09-917-376-3 (1-740)	X CB898982 (1-803)

Qy	407	AsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValIysGly	426
Db	67	CACGATTTCCCAACATCGGCACGCGCAAAATGTGTCAATCCAAATCAGTCGCGACAGCGC	126
Qy	427	LeuGluGluThrAlaValAsnAspLeuIleSerProProSerGlyAlaProLeuIleSer	446
Db	127	ATCGAGGAATTCCTCGTCCAGGACCTGGCTCTGCACCCGCGGAGGAGGACGTAATGGCC	186
Qy	447	AlaLeuGlyAspLeuGlyGlyPheThrHisAla	463
Db	187	GCAGTCGGAGACGACAACGGCTTCACCTTTGCCAGCAGAAACGACTCTGGGACATCGCGC	246
Qy	464	SerThiIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeu	483
Db	247	CAGACGGTCTGGGCAACGCCACATCGACGGCCACCTCGACGAGCGTCGACTACGCCGGGAAC	306
Qy	484	AsnProSerIleIleValArgAlaGlySerPheAsnProSerSerGlnProAsnAspArg	503
Db	307	TCGGTCAAGAGCGTGCCTCGCGTCGGCAACACGCCCGGCACGCAACAG-----	354
Qy	504	HisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGly	523
Db	355	---GTGCCCATCTCTCCGACGGCGCGCGAGCTGGAGCATCGACTACGCGCGCCGACACG	411
Qy	524	ValThrThrGlyGlyThrValAlaIleAspAlaAspGlySerArgPheValTrpAlaPro	543
Db	412	TCCATGAACCGCGGCACGGTGGCGCTATTTCGGCCGACGCGCACGATCTCTCTGGTCCGACC	471
Qy	544	GlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSer	563
Db	472	GCCTCGTCCGCG-----GTGCGAGCGCTCGCAGTTCCAGGGCAGCTTTTGCTTCCGTC	522
Qy	564	GlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyr	583
Db	523	TCGAGCTCGCCGGCGCGCGCTCATCGCTCGGACAAGAAGACCAACAGCGCTCTTCTAC	582
Qy	584	AlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProVal	603
Db	583	GCCGGCTCCGGATCGACCTTTTACGTTCAGCAAGGACACCGCGCAGCGAGCTTC-----	633
Qy	604	AlaAlaGlyLeuProSerSerGlyAlaValGlyValMet-----PheHisAlaValPro	621
Db	634	ACGCGCGGG---CCCAAGCTGGGCGAGCGCAGGAGCATCCGGGATATCGTCTGCCACCGC	690
Qy	622	GlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsnGly	641
Db	691	ACCACCGCGGCGAGTTGTATGTCTCGACCGCAGCTCGGCATATTTCGGCTCCACAGACTCG	750
Qy	642	GlySerSerTrpSerAlaIleThr	649
Db	751	GGCAGCAGCTTTGGCCAAAGTCTCC	774

RESULT 10	CF866388	738 bp	linear	EST 31-OCT-2003
LOCUS	CF866388			
DEFINITION	trico06xe08.b1 T.reesei mycelial culture, version 6 October 2003			
	Hypocrea jecorina cDNA clone trico06xe08, mRNA sequence.			
ACCESSION	CF866388			
VERSION	CF866388.1			
KEYWORDS	GI:38121014			
SOURCE	EST.			
	Hypocrea jecorina (anamorph: Trichoderma reesei)			

**FEATURES**  
**source**



US-09-917-376-3 (1-740) x CB896354 (1-794)

Qy 422 ProMetVallys---GlyLeuGluGluThrAlaValaAsnAspLeuIleSerProProSer 440  
 Db 58 CCACGCTCCGAGCGCATCGAGGATTCCTCGTCAGGACCTGGCTTCGACCCGGC 117  
 Qy 441 GlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAla----- 457  
 Db 118 GGAAGCGAGCTATTGGCGGAGTCGAGACGACCAACGCTTCACCTTTGCCAGCAGAAAC 177  
 Qy 458 AspValThrAlaValProSerThrIlePheThrSerProValPheThrGlyThrSer 477  
 Db 178 GACCTCGGACATCGCGCAGACGCTTGGCAGCGCCACATGGGCCCATCGACGAGC 237  
 Qy 478 ValAspTyrAlaGluLeuAsnProSerIleIleValAlaGlySerPheAspProSer 497  
 Db 238 GTCAGCTACCGCGGAACTCGTCAAGAGCGTCTCGCGTGGCGCAACACCGCGCGCAGC 297  
 Qy 498 SerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGln 517  
 Db 298 CAACAG-----GTGGCCATCTCGTCCGACGCGCGCGCGAGCATC 342  
 Qy 518 GlySerGluProGlyGlyValThrGlyGlyThrValAlaAlaSerAlaAspGlySer 537  
 Db 343 GACTACGCGCGCGACACGTCATGAACGCGCGGACGCTGGCTATTGGCGCGACGCGAC 402  
 Qy 538 ArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGly 557  
 Db 403 ACGATCTCTGGTCCGACCGCTCGTCGCG-----GTGCAGCGCTCGAGTTCCAG 453  
 Qy 558 AsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgVal 577  
 Db 454 GGCAGCTTTCGCTCGCTCGAGCTCCCGCGCGCGCTCATCGCTCGGACAGAG 513  
 Qy 578 AsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGly 597  
 Db 514 ACCAACAGCGCTCTTACTACGCGCGCTCGGATCGACCTTTTACGTCAGCAGACGACCGCGC 573  
 Qy 598 ValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMet--- 616  
 Db 574 AGCAGCTTC-----ACGCGCGG---CCCAAGCTGGCGCAGCGGACGATCCGG 621  
 Qy 617 ---PheHisAlaValProGlyLysGlyGlyAspLeuTrpLeuAlaAlaSerSerGlyLeu 635  
 Db 622 GATATCGCTCTACCCGACACCGCGCGGACGTTGTATGTCTCGACGCGAGCTCGGATA 681  
 Qy 636 TyrHisSerThrAsnGlySerSerTrpSerAlaIle---ThrGlyValSerSerAla 654  
 Db 682 TTCGCTCCACAGACTCCGCGGACGACCTTTGGCCAAAGTCTCCACCGCGCTGACCAACACC 741  
 Qy 655 ValAsnValGlyPheGlyLysSerAlaProGlySerSerTyr 668  
 Db 742 TACCAGATCGCTCGGTGTGGGTCTCA---GGCTCGAACTGG 780

RESULT 12  
 LOCUS BF072664 546 bp mRNA linear EST 18-OCT-2000  
 DEFINITION NCSM3H73 Subtracted Mycelial Neurospora crassa cDNA clone SM3H7 5' similar to avicelase III, Aspergillus aculeatus, mRNA sequence.  
 ACCSSION BF072664  
 VERSION BF072664.1 GI:10866169  
 KEYWORDS EST.  
 SOURCE Neurospora crassa  
 ORGANISM Neurospora crassa  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.  
 1 (bases 1 to 546)  
 REFERENCE  
 AUTHORS Leonard, P.M., Mitchell, J., Amijo, A.M., Bean, L., Dolan, P.L.,  
 Cushing, T., Errett, A., Fleharty, M., Gorman, M., Judson, K.,  
 Miller, R., Ortega, J., Pavlova, I., Perea, J., Todisco, S.,  
 Trujillo, R., Valentine, J., Wells, A., Werner-Washburne, M., Yazzie, S.  
 and Natvig, D.O.

TITLE Expressed sequences from conidial, mycelial, and sexual stages of  
 Neurospora crassa  
 JOURNAL Fungal Genet. Biol. 21, 348-363 (1997)  
 MEDLINE 97435549  
 PUBMED 9290248  
 COMMENT Contact: Natvig, D.O./Nelson, M.A.  
 Department of Biology  
 University of New Mexico  
 Castetter Hall, Albuquerque, NM 87131, USA  
 Tel: 505 277 3411  
 Fax: 505 277 0304  
 Email: ngp@biology.unm.edu.  
 Location/Qualifiers

FEATURES  
source

1..546  
 /organism="Neurospora crassa"  
 /mol\_type="mRNA"  
 /strain="74-OR23-IV A (FGSC 2489)"  
 /db\_xref="taxon:5141"  
 /clone="SM3H7"  
 /sex="Mating type A"  
 /tissue\_type="Mycelium"  
 /dev\_stage="Mycelium"  
 /lab\_host="E. coli"  
 /clone\_lib="Subtracted Mycelial"  
 /note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:  
 XhoI; 2% sucrose for 24 hours. cDNA directionally cloned  
 into pBluescript SK(-) using the Uni-ZAP XR vector system  
 (Stratagene, La Jolla, CA). Previously identified highly  
 expressed clones were subtracted from this library."

## ORIGIN

Alignment Scores:  
 Pred. No.: 3 29e-17 Length: 546  
 Score: 314.50 Matches: 72  
 Percent Similarity: 54.26% Conservative: 30  
 Best Local Similarity: 38.30% Mismatches: 75  
 Query Match: 7.79% Indels: 11  
 DB: 10 Gaps: 3  
 US-09-917-376-3 (1-740) x BF072664 (1-546)

Qy 358 SerAlaGluProTrpLeuThrPheGlyValGlnProAsnProProValProSerProLys 377  
 Db 5 TCGGCACGAGCTGGATCGAGACTGGCTCTTTTCAG-----GATACCAAGCAC 55  
 Qy 378 LeuGlyTrpMetAspGluAlaMetAlaIleAspPropheAsnSerAspArgMetLeuTyr 397  
 Db 56 CTCGCTGGATGATCGAATCCCTCGAGATCAACCCCTCGACAGCGACCATTCGCTCTAC 115  
 Qy 398 GlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGln 417  
 Db 116 GCCACCGTCTCAGTGTTCACGGCGGCGGACGACCTGACCAAGTGGGACACCGTCCACAC 175  
 Qy 418 IleHisIleAlaProMetValLysGlyLeuGluGluThrAlaValAsnAspLeuIleSer 437  
 Db 176 GTGACCATCCAGTCTCTGGCGCTTGGCATCGAAGAAATGGCTCTCGCTCTCGCTCC 235  
 Qy 438 ProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThr----- 455  
 Db 236 GCCCGCGGTGGCTCGAGCTTCTCGCGCGCTCGGTGACGACTGCGCGCTTCACCTTCAAG 295  
 Qy 456 ---HisAlaAspValThrAlaValProSerThrIlePheThrSerProValPheThr 474  
 Db 296 TCCAGCTCCGACCTCGGCACCTCTCCCAAGACCCCTGATGATGACCCCGCGGCGGCGCAGC 355  
 Qy 475 GlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPhe 494  
 Db 356 TCCGCGGAGCTCGACTAGCCGCGCAACAGCCGCGCAAGCTGGTGGCATCGGCTCGGCGC 415  
 Qy 495 AspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsn 514  
 Db 416 TCCGCGCGCGCAGCAG-----GTGGCGGTATCTCTCGGACGCGCGCGCTCG 460

QY 515 TrpPheGlnGlySerGluProGlyGlyValThrThrThrGlyGlyThrValAlaAlaSerAla 534  
Db 461 TGGCATGCTCAACAGCGGACGACACACCAAGAGCAGCGGCTGCTCTCTGCTGCC 520  
QY 535 AspGlySerArgPheValTIPAla 542  
Db 521 GACGCCGACCATCGTCTGTCT 544

RESULT 13  
BU639045 389 bp mRNA linear EST 06-MAY-2003  
LOCUS mcgw011xd18.f RCW Lambda Zap Express Library Magnaporthe grisea  
DEFINITION cDNA clone mcgw011xd18 5', mRNA sequence.  
ACCESSION BU639045  
VERSION BU639045.2 GI:30397224  
KEYWORDS EST.  
SOURCE Magnaporthe grisea (anamorph: Pyricularia grisea)  
ORGANISM Magnaporthe grisea  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Sordariomycetes incertae sedis; Magnaportheaceae; Magnaporthe.  
REFERENCE 1 (bases 1 to 389)  
AUTHORS Ebbole,D.J., Yuan,J., Thomas,T.L., Bobrowicz,P., Lu,G.,  
Bhatteirai,K. and Dean,R.A.  
TITLE Expressed sequence tags from the rice blast fungus, Magnaporthe  
grisea  
JOURNAL Unpublished (2002)  
COMMENT On Sep 30, 2002 this sequence version replaced gi:23351371.  
Contact: Ebbole DJ  
Department of Plant Pathology & Microbiology  
Texas A&M University  
Peterson Bldg, MS2132, College Station, TX 77843-2132, USA  
Tel: 979 845 4831  
Fax: 979 845 6483  
Email: d-ebbole@amu.edu  
Chromatogram file of this sequence is available, see contact  
person; Best nr hit (April. 22, 2003) gb|EAA29333.1| Hypoetical  
protein (Neurospora crassa) 206 4e-53  
PCR Primers  
FORWARD: T3 primer  
BACKWARD: T7 primer  
Plate: mcgw011 row: D column: 18  
Seq primer: T3.

FEATURES  
source  
1..389 Location/Qualifiers  
/organism="Magnaporthe grisea"  
/mol\_type="mRNA"  
/strain="CP987"  
/db\_xref="taxon:148305"  
/clone="mcgw011xd18"  
/sex="Mat1-1 hermaphrodite"  
/tissue type="Mycelium"  
/dev stage="Day 5 post-inoculation"  
/clone\_lib="RCW Lambda Zap Express Library"  
/notes="Vector: pBluescript excised from Lambda Zap  
Express; Site 1: EcoRI; Site 2: XhoI; Day 5  
post-inoculation MRNAs prepared from Magnaporthe grisea  
grown at 23C in the dark with constant gyratory shaking  
100 rpm in Vogel's minimal medium containing 0.5% isolated  
rice cell walls as the sole carbon source. Library  
provided by Sheng-Cheng Wu. Sequences were processed by  
one of two methods. Where a full-length alignment to the  
M. grisea genome sequence was available, the EST sequence  
was trimmed according to the alignment, otherwise sequence  
quality was assessed using phredPhrap version 991019 and  
trimmed according to phd files (0.05) and for vector  
seqs."

ORIGIN  
Alignment Scores:  
Pred. No.: 9.49e-17 Length: 389  
Score: 306.50 Matches: 55  
Percent Similarity: 66.67% Conservative: 27  
Best Local Similarity: 44.72% Mismatches: 38

Query Match: 7.59% Indels: 3  
DB: 13 Gaps: 1  
US-09-917-376-3 (1-740) x BU639045 (1-389)

QY 334 GlyAlaThrThrThrArgIleTrpAspTrpThrSerTyPProAsnArgSerLeuArgTyr 353  
Db 5 GGGGCTACATGTCGAGACTTTGGGAGTGGAGCGATATCCGACATGAACCACTACTAC 64

QY 354 ValLeuAspIleSerAlaGluProTTPLeuThrPheGlyValGlnProAsnProProVal 373  
Db 65 TCGATCAACACAGACAGCGCCCTGGATAGTCTGGC-----CATCTTCGAGG 115

QY 374 ProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAsp 393  
Db 116 GATTGCGAGCGCTTGGTTGGATGATCGAGGCCCTTGAATCGACCCCTGTCGATCCTGAC 175

QY 394 ArgMetLeuTyGlyThrGlyAlaThrLeuTyAlaThrAsnAspLeuThrLysTrpAsp 413  
Db 176 CATGGCTTTACGGAACCTGCTCTCGATCTTTGGGGGACATGACTTTGACCAATGGGAC 235

QY 414 SerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluThrAlaValAsn 433  
Db 236 ACGGTTCAACATATCGATTCACTTTGGCCGATGGTGTGGAAGAACTGCGATTCTG 295

QY 434 AspLeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGly 453  
Db 296 GAGATGGCTTACGCGCCGCGAGGCTCCGAATCTGCGCGCAATGGAGACATTACTGGA 355

QY 454 PheThrHis 456  
Db 356 TTCACATAC 364

RESULT 14  
CF875815 713 bp mRNA linear EST 31-OCT-2003  
LOCUS trico39xk06.b1 T.reesei mycelial culture, Version 6 October 2003  
DEFINITION Hypocrea jecorina cDNA clone trico39xk06, mRNA sequence.  
ACCESSION CF875815  
VERSION CF875815.1 GI:38130497  
KEYWORDS EST.  
SOURCE Hypocrea jecorina (anamorph: Trichoderma reesei)  
ORGANISM Hypocrea jecorina  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.  
REFERENCE 1 (bases 1 to 713)  
AUTHORS Diener,S.E., Dankmeyer,L., Dunn-Coleman,N., Houfek,T.D.,  
Mitchell,T.K., van Solingen,P., Teunissen,P.J.M., Ward,M. and  
Dean,R.A.  
TITLE Analysis of the protein processing and secretion pathways in a  
Trichoderma reesei EST dataset  
JOURNAL Unpublished (2003)  
COMMENT Contact: Ralph A. Dean  
Fungal Genomics Laboratory  
North Carolina State University  
Campus Box 7251, Raleigh, NC 27695, USA  
Tel: 919-513-0020  
Fax: 919-513-0024  
Email: ralph.dean@ncsu.edu  
Seq primer: Tr-F1 primer.  
Location/Qualifiers  
1..713  
/organism="Hypocrea jecorina"  
/mol\_type="mRNA"  
/strain="QM6a"  
/db\_xref="taxon:51453"  
/clone="trico39xk06"  
/dev stage="mycelia"  
/clone\_lib="T.reesei mycelial culture, Version 6 October  
2003"  
/note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial  
culture grown from 24 hrs to 6 days with varying Carbon  
and Nitrogen sources and concentrations."

FEATURES  
source  
1..713  
/organism="Hypocrea jecorina"  
/mol\_type="mRNA"  
/strain="QM6a"  
/db\_xref="taxon:51453"  
/clone="trico39xk06"  
/dev stage="mycelia"  
/clone\_lib="T.reesei mycelial culture, Version 6 October  
2003"  
/note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial  
culture grown from 24 hrs to 6 days with varying Carbon  
and Nitrogen sources and concentrations."

```

ORIGIN
Alignment Scores:
Pred. No.: 1.42e-14 Length: 713
Score: 286.50 Matches: 81
Percent Similarity: 48.79% Conservative: 40
Best Local Similarity: 32.66% Mismatches: 106
Query Match: 7.10% Indels: 21
DB: 14 Gaps: 10

US-09-917-376-3 (1-740) x CF875815 (1-713)
Qy 458 AspValThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSer 477
Db 10 GACCTCGGTATCATCGCGAGTGTGTTCTGGCAACGCCCATCGGCACCTCGACGAGC 69
Qy 478 ValAspTyrAlaGluLeuAsnProSerIleValArgAlaGlySerPheAspProSer 497
Db 70 GTCAGTACGCGGGAACCTCGTCAAGAGCGTGTGCGGTGCGCAACACCGCGCGGACG 129
Qy 498 SerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGln 517
Db 130 CAACAG-----GTGGCCATCTCTCGTCGAGCGCGCGGCGGAGCATC 174
Qy 518 GlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySer 537
Db 175 GACTACGCGCGGACAGCATCATGAACGCGGCGGCGGCGGCGGCTATTTCGCGGACGCGGAC 234
Qy 538 ArgPheValTyrAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGly 557
Db 235 ACGATCTCTGTGTCAGCGCGCTCGTCGCGC-----GTGACGCGCTCGAGTTCCAG 285
Qy 558 AsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgVal 577
Db 286 GGCAGCTTGTCTCGTCTGAGCGCTGCCCGCGGCGGCGGCTATCGCTCGGACAGAG 345
Qy 578 AsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGly 597
Db 346 ACCAACAGCGTCTTCTACGCGCGGCTCGGATCGACCTTTTACGTCAGCAAGACACCGGC 405
Qy 598 ValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMet--- 616
Db 406 AGCAGCTTC-----ACGCGCGGG---CCCAAGCTGGCGGCGGCGGAGCATCGG 453
Qy 617 ---PheHisAlaValProGlyLysGlyLysPheThrLeuAlaAlaSerSerGlyLeu 635
Db 454 GATATCGCTGTCTACCCGACACCGCGGCGGCGGCTGTGTATGTCTCGACGAGTCGGGATA 513
Qy 636 TyrHisSerThrAsnGlyGlySerSerTrpSerAlaIle---ThrGlyValSerSerAla 654
Db 514 TTCGCTCCACAGACTCGGGGACAGACCTTGGCCAGTCTCCACCGCCCTGACCCACACC 573
Qy 655 ValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValVal 674
Db 574 TACCAGATCGCCCTGGGTGTGGGTCA---GGCTCGAACTGG---AACCTGATGCTTC 627
Qy 675 GlyThrIleGlyGlyValThrGlyAla-----TyrArgSerAspAspCysGlyThrThr 692
Db 628 GGCACC-----GGCCGCTAGGGGCTGCGCTCTACGCCAGTGAGAGACGCGGCGCTCC 681
Qy 693 TrpValLeuIleAsnAspAspGln 700
Db 682 TGGACGACATTTAGGGCTCCAG 705

RESULT 15
CB904767
LOCUS
DEFINITION
trio039xk06 T.reesei mycelial culture, Version 3 april Hypocrea
jecorina cDNA clone trio039xk06, mRNA sequence.
ACCESSION
CB904767.1 GI:30119425
VERSION
CB904767.1
KEYWORDS
Hypocrea jecorina (anamorph: Trichoderma reesei)

```

```

ORGANISM Hypocrea jecorina
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
REFERENCE
1 (bases 1 to 782)
AUTHORS
Foreman, P.K., Brown, D.E., Dankmeyer, L., Dean, R., Diener, S.,
Dunn-Coleman, N.S., Goedegebuur, F., Houfek, T.D., England, G.J.,
Kelley, A.S., Meerman, H.J., Mitchell, T., Mitchinson, C.,
Olivares, H.A., Teunissen, P.J., Yao, J.J. and Ward, M.
TITLE
Transcriptional regulation of biomass-degrading enzymes in the
filamentous fungus Trichoderma reesei
J. Biol. Chem. 278 (34), 31988-31997 (2003)
JOURNAL
2803314
MEDLINE
12788920
PUBMED
COMMENT
Contact: Pamela K. Foreman
925 Page Mill Road, Palo Alto, CA 94304, USA
Tel: (650) 846-7635
Fax: (650) 621-7817
Email: Pforeman@genencor.com
Seq primer: LT-F1 primer.
FEATURES
Location/Qualifiers
1..782
source
/organism="Hypocrea jecorina"
/mol_type="mRNA"
/strain="QM6a"
/db_xref="taxon:51453"
/clone="trio039xk06"
/dev_stage="mycelia"
/clone_lib="T.reesei mycelial culture, version 3 april"
/note="Vector: pREP3; Site 1: Not I/Sal I; Mycelial
culture grown from 24 hrs to 6 days with varying Carbon
and Nitrogen sources and concentrations."
ORIGIN
Alignment Scores:
Pred. No.: 1.65e-14 Length: 782
Score: 286.50 Matches: 81
Percent Similarity: 48.79% Conservative: 40
Best Local Similarity: 32.66% Mismatches: 106
Query Match: 7.10% Indels: 21
DB: 14 Gaps: 10

US-09-917-376-3 (1-740) x CB904767 (1-782)
Qy 458 AspValThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSer 477
Db 79 GACCTCGGTATCATCGCGAGTGTGTTCTGGCAACGCCCATCGGCACCTCGACGAGC 138
Qy 478 ValAspTyrAlaGluLeuAsnProSerIleValArgAlaGlySerPheAspProSer 497
Db 139 GTCAGTACGCGGGAACCTCGTCAAGAGCGTGTGCGGTGCGCAACACCGCGGACG 198
Qy 498 SerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGln 517
Db 199 CAACAG-----GTGGCCATCTCTCGTCGAGCGGCGGCGGCGGAGCATC 243
Qy 518 GlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySer 537
Db 244 GACTACGCGCGGCGACACGTCCTCAACGCGGCGGCGGCTATTCGCGCGGCGGAC 303
Qy 538 ArgPheValTyrAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGly 557
Db 304 ACGATCTCTGTGTCAGCCGCTCTGTCGCGC-----GTGACGGCTCGGAGTTCCAG 354
Qy 558 AsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgVal 577
Db 355 GGCAGCTTGTCTCGTCTCGAGCGTGTGCGGCGGCGGCGGCTATTCGCGCGGCGGAC 414
Qy 578 AsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGly 597
Db 415 ACCAACAGCGTCTTCTACGCGGCTCTCGGATCGACCTTTTACGTCTCAGCAAGACACCGGC 474
Qy 598 ValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMet--- 616

```

475	Db	AGCAGCTTC	:::			:::	:::	:::	522
617	Qy	--PheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSergGlyLeu				:::	:::	:::	635
523	Db	GATATCGCTGCTACCCGACACCGCGGGCAGCTTGTAATGCTCGACCGACGTCGCGCAT				:::	:::	:::	582
636	Qy	TyrHisSerThrAsnGlyGlySerSerTrpSerAlaIle				:::	:::	:::	654
583	Db	TTCCGCTCCACAGACTCGGGCAGCACACTTGGCCCAAGTCTCCACCGCCCTGACCAACACC				:::	:::	:::	642
655	Qy	ValAsnValGlyPheGlyLysSerAlaProGlySerSerSerTyrProAlaValPheValVal				:::	:::	:::	674
643	Db	TACCAGATCGCCCTGGGTGTGGGTCA				:::	:::	:::	696
675	Qy	GlyThrIleGlyGlyValThrGlyAla				:::	:::	:::	692
697	Db	GGCACC				:::	:::	:::	750
693	Qy	TrpValLeuIleAsnAspGln				:::	:::	:::	700
751	Db	TGGACGGACATTCAGGGCTCCAG				:::	:::	:::	774

Search completed: May 12, 2004, 11:38:55  
Job time : 6964.51 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 12, 2004, 04:43:36 ; Search time 184.798 Seconds  
(without alignments)  
2222.227 Million cell updates/sec

Title: US-09-917-376-3

Perfect score: 4036

Sequence: 1 ATTQPTWNSNAVGCGFVD.....YGTNGRGIVYDGGAPSG 740

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09917376/runat\_11052004\_114536\_28265/app\_query.fasta\_1.1.1429  
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -IOFCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09917376 @cgn\_1\_116 @runat\_11052004\_114536\_28265 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=10 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:

- 1: /cgn2\_6/prodata/2/ina/5A COMB.seq.\*
- 2: /cgn2\_6/prodata/2/ina/5B COMB.seq.\*
- 3: /cgn2\_6/prodata/2/ina/6A COMB.seq.\*
- 4: /cgn2\_6/prodata/2/ina/6B COMB.seq.\*
- 5: /cgn2\_6/prodata/2/ina/6C COMB.seq.\*
- 6: /cgn2\_6/prodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	227.5	5.6	3129	US-09-252-991A-13873	Sequence 13873, A
C 2	227.5	5.6	8211	US-09-252-991A-13656	Sequence 13656, A
C 3	203	5.0	4188	US-09-252-991A-13774	Sequence 13774, A
C 4	201.5	5.0	4131	US-09-252-991A-13773	Sequence 13773, A
C 5	192	4.8	4411529	US-09-103-840A-1	Sequence 1, Appli
C 6	184	4.6	4403765	US-09-103-840A-2	Sequence 2, Appli
C 7	181.5	4.5	11679	US-09-328-352-1377	Sequence 39, Appl
C 8	170.5	4.2	25165	US-09-453-702B-39	Sequence 2, Appli
C 9	170	4.2	4403765	US-09-103-840A-2	Sequence 2, Appli
C 10	169	4.2	4411529	US-09-103-840A-1	Sequence 1, Appli
C 11	164	4.1	2319	US-09-252-991A-13875	Sequence 13875, A
C 12	163	4.0	5741	US-07-706-699-4	Sequence 4, Appli

13	163	4.0	5741	1	US-07-998-931-4	Sequence 4, Appli
14	160.5	4.0	47981	4	US-09-679-279-1	Sequence 1, Appli
15	159.5	4.0	4647	4	US-09-252-991A-5730	Sequence 5730, Ap
16	159.5	4.0	10419	4	US-09-408-020-3	Sequence 3, Appli
17	159.5	4.0	42432	4	US-09-408-020-2	Sequence 2, Appli
18	157.5	3.9	2067	4	US-09-489-039A-2642	Sequence 2642, Ap
19	156.5	3.9	1860	4	US-09-252-991A-9781	Sequence 9781, Ap
20	155	3.8	2235	1	US-08-418-782-1	Sequence 1, Appli
21	155	3.8	2235	2	US-08-228-662-1	Sequence 1, Appli
22	155	3.8	2235	2	US-08-852-219-1	Sequence 1, Appli
23	154	3.8	3900	4	US-09-023-655-1420	Sequence 1420, Ap
24	153	3.8	2625	6	5457037-4	Patent No. 5457037
25	153	3.8	3336	6	5457037-1	Patent No. 5457037
26	153	3.8	18034	4	US-09-266-965-75	Sequence 75, Appl
27	152.5	3.8	6416	3	US-09-136-574A-2	Sequence 2, Appli
28	151	3.7	1989	4	US-09-252-991A-15371	Sequence 15371, A
29	151	3.7	2637	4	US-09-252-991A-15371	Sequence 15403, A
30	150.5	3.7	2058	4	US-09-252-991A-15403	Sequence 15462, A
31	149.5	3.7	2253	6	5457037-2	Patent No. 5457037
32	149	3.7	2852	4	US-09-056-556-203	Sequence 203, App
33	149	3.7	2852	4	US-09-072-596-198	Sequence 198, App
34	149	3.7	2852	4	US-09-072-967-203	Sequence 203, App
35	149	3.7	3120	4	US-09-169-768-29	Sequence 29, Appl
36	147.5	3.7	44377	2	US-08-804-227C-7	Sequence 7, Appli
37	147.5	3.7	44377	2	US-08-804-198-1	Sequence 1, Appli
38	146	3.6	87563	4	US-09-453-702B-57	Sequence 57, Appl
39	145	3.6	2163	3	US-09-296-284-22	Sequence 22, Appl
40	145	3.6	2185	1	US-08-173-508-3	Sequence 3, Appli
41	145	3.6	2185	2	US-08-265-310-3	Sequence 3, Appli
42	145	3.6	2185	3	US-08-951-742-3	Sequence 3, Appli
43	145	3.6	2265	3	US-09-296-284-1	Sequence 1, Appli
44	145	3.6	4830	3	US-09-296-284-7	Sequence 7, Appli
45	144.5	3.6	1164	4	US-09-252-991A-5391	Sequence 5391, Ap

ALIGNMENTS

RESULT 1

US-09-252-991A-13873/c  
; Sequence 13873, Application US/09252991A  
; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 13873

; LENGTH: 3129

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-13873

Alignment Scores:	7,76e-11	Length:	3129
Pred. No.:	227.50	Matches:	218
Score:	31.30%	Conservative:	90
Percent Similarity:	22.15%	Mismatches:	360
Best Local Similarity:	5.64%	Indels:	319
Query Match:	4	Gaps:	52
DB:			

US-09-917-376-3 (1-740) x US-09-252-991A-13873 (1-3129)

Qy 20 AspGlyIleValPheAenGluGlyAla---ProGlyIleLeuTyValArgThrAspIle 38

Db 2993 AACGGCGTCGTCATCAGCGGACCGCGGAGCGCGTCCACCGTACCGTCACCGATGCC 2934

Qy 39 GlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpVal 58  
Db 2933 GGC-----GGCAACCCGATGAGCGCAGGTCAACCGCGAC 2901  
Qy 59 GlyTyrAsnAsnTrpGlyTyr-----AsnGly-----ValVal 69  
Db 2900 GGCAGCGGCACTGGAGCTTACCGCGGGGACCGCGCGGCAACGGCAGGTGATGTC 2841  
Qy 70 SerIleAlaAlaAspProIle---AsnThrAsnLysValTrpAla-----Ala 84  
Db 2840 GCCAGCGCCACCGACCGCGCAATACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2781  
Qy 85 ValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu-----100  
Db 2780 GTGGCG 2721  
Qy 101 -----ArgSer 102  
Db 2720 GAGCG 2661  
Qy 103 SerAspGlnGlyAlaThrTrpGlnIleThrPro---LeuProPheLysLeuGly-----119  
Db 2660 GCCGAGCG 2601  
Qy 120 -----GlyAsnMetProGlyArgGlyMetGlyGluArg 130  
Db 2600 GTCAACCG 2541  
Qy 131 LeuAlaValAspProAsn-----AsnAspAsnIleLeuTyrPhe 143  
Db 2540 GAGCGGTGGCG 2481  
Qy 144 GlyAlaProSerGlyLysGlyLeu-----Trp 152  
Db 2480 ACCG 2421  
Qy 153 ArgSerThrAspSerGlyAlaThrTrp-----SerGlnMetThrAsnPhe 167  
Db 2420 ACCACCG 2364  
Qy 168 ProAspValGlyThrTyrIle---AlaAsnProThrAspThrThrGlyTyrGlnSerAsp 186  
Db 2363 -----GGCACCGGTGGTCAACGTCGACCGCGCGCGCGCGCGCGCGCGCGCG 2316  
Qy 187 IleGlnGlyValValTrpValAlaPheAspLysSerSerSerSerLeuGlyGln-----204  
Db 2315 -----GCTCCCGCTACCGAGCGGTGGATTCCTGCTGCGTGGATCCCGCGAGTGGAT 2262  
Qy 205 ---AlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPhe-----221  
Db 2261 CCGAGCAACGGTTCCGGTGATCAGCGGCAACCGCGCGCGCGCGCGCGCGCGCGCG 2202  
Qy 222 -----TrpSerArgAspGlyGlyAlaThrTrpGlnAla 232  
Db 2201 GATGGCAACGGCAACCGGATTTGCCAGGTCAACCGCGCGCGCGCGCGCGCGCGCG 2142  
Qy 233 ValProGlyAlaPro-----237  
Db 2141 ACTCCAGGCATCCCGTCCGGATGGCAGCGGTGTCACCGTGGTGGCGCGCGCGCGCG 2082  
Qy 238 -----ThrGlyPheIleProHisLysGlyVal 246  
Db 2081 AATGTGCAGATCGCGCGCGGTGATCACTGTGATGGCGTGGCGCGCGCGCGCGCG 2022  
Qy 247 PheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyr---265  
Db 2021 ATCGATCCGAGCAACCGCGCACCGAGATGAAGCGGTACCGCGGAGCGCGCGCGCG 1962  
Qy 266 -----AspGlySerSerGlyAspValTrpLysPheSerValThr---SerGlyThrTrp 282  
Db 1961 CTCACG 1902  
Qy 283 ThrArgIleSerProValPro-----SerThrAsp 292

Db 1301 ACGTTACCCCGCGCACCCCGCTGGCCCAACCGCACCGTGCATCAACCGCGTGGCCGAC 1842  
Qy 293 ThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisPro-----310  
Db 1841 CCGCGCGCAATACACGCGGTCCCGCGCAGCGTCAACCGTGCATGCCATGCCCGCGCGCG 1782  
Qy 311 -----AsnThrIleMetValAlaThrGlnIleSerTrpTrpProAsp 324  
Db 1781 CCGGTGATCAATCCGAGCAACGAGTCTCATCGCGGTACCGCGGAACCGCGCGCGCG 1722  
Qy 325 ThrIleIlePheArgSerThrAspGlyGly-----AlaThrTrpThr 338  
Db 1721 GTGATCTCTC-----ACCGAGCGCAACCGCAACCGCATCGCGCAGTCAACCGCGCG 1671  
Qy 339 ArgIleTrpAspTrpThr-SerTyrProAsnArg-----SerIle 351  
Db 1670 GCGAGCGCAAGTGGGCTTTCACGCCCGCCACCGCTGGCGCAATGGCACGGTGCATCAAT 1611  
Qy 351 uArgTyrValIleuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnPr 371  
Db 1610 GCGCTGG-----CCGAGGACCGCGCGCGCAACACAGCAGTC-----CC 1572  
Qy 371 oProValProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspPropheAs 391  
Db 1571 ACCAGCGCCA---CCGTGCACTCGTGGCGCGCA-GCCCCGGTGCATCGATCCGAGCAA 1516  
Qy 391 nSerAspArgMetLeuTyrGlyThr-----GlyValThrLeuTyrAlaThrAsnAs 408  
Db 1515 CGGTAGC---GTGATCGCGGTACCGCGAGCGTGGTCCACGGTGCCTCCTCACC-----1464  
Qy 408 pLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuG 428  
Db 1463 -----GACGGAACGGCAAC-----CCGATCGCGCGCGTCAACCGC 1429  
Qy 428 uGluThrAlaValAsnAspLeuIleSerProSerGlyAlaProLeu-----444  
Db 1428 CGATGGCAGCGGCACTGGAGCTTCACGCC-----GGCACCGCGCTGTCCAATGGCAC 1375  
Qy 445 ---IleSerAlaLeuGly---AspLeuGlyGlyPheThrHisAlaAspVal---ThrAl 461  
Db 1374 GTGTGTCATGCGGTGGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1315  
Qy 461 aValProSerThrIlePheThrSerProValPhe-----ThrThrGlyThrSerVal--478  
Db 1314 GTGTGACTCGTGGCG 1255  
Qy 479 -AspTyrAlaGluLeuAsnProSerIleValArgAlaGlySerPheAspProSerSe 498  
Db 1254 CCGTACCGCGGAGCGCGGTGGCGCGGTGATCTCTACCGATGGCGCGCGCAACCGCGTGG 1195  
Qy 498 rGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrp-----515  
Db 1194 CCAGGCC-----ACCGCGGTGGCAGCGCGCAACTGGAGTCTCAC 1156  
Qy 516 -----PheGlnGlySerGluProGlyG 523  
Db 1155 CCGCGGCGACCGCGTGGCGCAACGCGCATCAATCGGTGGCGCGCGCGCGCGCGCG 1096  
Qy 523 yValThrThrGlyGlyThr-----529  
Db 1095 CAATACCG 1036  
Qy 530 -----ValAlaAlaSerAlaAspGlySerArgPheValTrpAl 542  
Db 1035 CAACCCGAGCAACGCGCAGCGTGCATCGCGGTGTACCGCGGAACCGCGCGCGCGTGTCT 976  
Qy 542 aProGlyAspProGlyGlnProVal-----ValTyrAlaValGlyPheGlyAsnSerTr 560  
Db 975 CACG 919  
Qy 560 pAlaAlaSerGlnGlyValPro---AlaAsnAlaGlnIleArgSer-----574

Db 918 GAGCTTACGCGCCGACGCGCTGGCCCAACGGCTCGGTGATCAATGCGCTGGCCGACGA 859  
Qy 575 -----AspArgValAsnProLysTh 581  
Db 858 CGCCCGCGGCAACACACGCGCGCGCCGACGACCGGTGGACTCGGTAGCCCGGCCAC 799  
Qy 581 r----PheTyrAlaLeuSerAsnGlyThrPheTyrArgSer----- 593  
Db 798 CCGCGTGTCTCGATCCGAGCAACGGTACGGTGATCAGCGGTACCGCGGAAGCCGGGCCAC 739  
Qy 594 -----ThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeu-ProSerS 610  
Db 738 GGTGATCTCTACCGACGCGCGC-----GGCAACCCGATACGCGACGCGCCGATG 685  
Qy 610 erGlyAlaValGlyValMetPheHisAlaValProGlyLysGluGlyAspLeuTLeuA 630  
Db 684 CAGCGCAACTGGAGCTTCACTCCGGGACACA-CCGCTGACCAACGCGACGGTGATCAATG 626  
Qy 630 laAlaSerSerGlyLeuTyrHisSerThrAsnGly-----GlySerSerTrpSerAlaI 648  
Db 625 CGTGCGCCGACGACCGCGCGCAACACCGCGGTCCGCTCAGCACCAACAGTGGAGCGG 566  
Qy 648 leThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSer 668  
Db 565 TGGCCCCCGCCACACCGCGGTGATCGACCGAGCAATGGTGTCAAACTCAGCGCACCGCG 506  
Qy 668 yrProAlaValPheValValGlyThrIleGlyGly----- 679  
Db 505 AACCCGCGGTCCGGTGATCTCACCAGTGGCAATGGCAACCGGATCGGCCAGACCTCG 446  
Qy 680 -----ValT 681  
Db 445 CCGAGGTAGCGCAACTGGACCTTCACACCGGGACGCGGTGGCCACGCGCGTGG 386  
Qy 681 hrGlyAlaTyrArgSerAspAspCysGlyThrThrTrpValLeuIleAsnAspAspGlnH 701  
Db 385 TCAACGCGGTGGCCCGGACCGCGCGCGCAATACCAAG-----CGTCCGCGCAGC 336  
Qy 701 isGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgArg-Val 720  
Db 335 ACCACGCGTGGATACGCTGGCGCGCGCCACGCGCGGTGATCAATGCCCAACCGCGCGGTG 276  
Qy 721 TyrIleGlyThrAsn-----GlyArgGlyIleValTyrGlyAspIleGlyGlyAlaPro 738  
Db 275 ATCAGCGSACCGCGAGGTGCGCGCGCAAGTGTCTCACCAGCGGCAACCGCAACCGCG 216  
Qy 739 SerGly 740  
Db 215 ATCGCG 210

RESULT 2  
US-09-252-991A-13656  
; Sequence 13656, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 13656  
; LENGTH: 8211  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-13656  
Alignment Scores:

Pred. No.: 3,45e-10 Length: 8211  
Score: 227.50 Matches: 218  
Percent Similarity: 31.30% Conservative: 90  
Best Local Similarity: 22.15% Mismatches: 360  
Query Match: 5.64% Indels: 319  
DB: 52  
US-09-917-376-3 (1-740) x US-09-252-991A-13656 (1-8211)  
Qy 20 AspGlyIleValPheAsnGluGlyAla---ProGlyIleLeuTyrValA-gThrAspIle 38  
Db 3574 AACGGGTGTATCATAGCGCGCACCGCGCGGTGGCCACCGGTGACCTCACCAGTATGCC 3633  
Qy 39 GlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpVal 58  
Db 3634 GGC-----GGCAACCCGATAGCGGAGGTCAACGCCGAC 3666  
Qy 59 GlyTrpAsnAsnTrpGlyTyr-----AsnGly-----ValVal 69  
Db 3667 GGCAGCGGCAACTGGAGCTTCACCGCGGCGCACCGCGCGGCAACCGCACCGGTATCGTC 3726  
Qy 70 SerIleAlaAlaAspProIle---AsnThrAsnLysValTrpAla-----Ala 84  
Db 3727 GGCAGCGCACCGCACCGCGCAATACCGCGCGGCGCGCGCACCGGTGACGCG 3785  
Qy 85 ValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu----- 100  
Db 3787 GTGGCGCGCGCGCGCGGTGATCGATCCGAGCAACGCGCACCGCATCAGCGGCGCACCG 3846  
Qy 101 -----ArgSer 102  
Db 3847 GAGCGCGCGCGCAAGGTGATCTTCACCGACGCGCAACCGCAACCGATCGGCGAAACCA 3906  
Qy 103 SerAspGlnGlyAlaThrTrpGlnIleThrPro---LeuProPheLysLeuGly----- 119  
Db 3907 GCGCAGCGCAGCGCAACTGGACCTTCACCGCGCGCGCACCGCGTGGCCACCGCGCGGTG 3966  
Qy 120 -----GlyAsnMetProGlyArgGlyMetGlyGluArg 130  
Db 3967 GTCAACCGCGTGGCCCGGACCGCTTGGGGCAATACCGCGCGCGCGGAGGAGCACTACCGTG 4026  
Qy 131 LeuAlaValAspProAsn-----AsnAspAsnIleLeuTyrPhe 143  
Db 4027 GACGCGGTGGCGCGCAACACGCGCTGTGTCAATCCGAGCAACGCGCAACCTCTCAACCGT 4086  
Qy 144 GlyAlaProSerGlyLysGlyLeu-----Trp 152  
Db 4087 ACCGCGCGCGCGCGGACGACCGTTCACCGACGCGCAACCGCAACCGATCGGCAG 4146  
Qy 153 ArgSerThrAspSerGlyAlaThrTrp-----SerGlnMetThrAsnPhe 167  
Db 4147 ACCACCGCGATGGCGAGCGCAACTGGAGCTTCACGCGCGCGTCCGCACTACCAACAC--- 4203  
Qy 168 ProAspValGlyThrTyrIle---AlaAsnProThrAspThrThrGlyTyrGlnSerAsp 186  
Db 4204 -----GGCACCGTGGTCAACGTCGCGAGCGACGCGCGCGGCAATACCAAGC--- 4251  
Qy 187 IleGlnGlyValValTrpValAlaPheAspLysSerSerSerSerLeuGlyGln----- 204  
Db 4252 -----GCTCCCGCTACACGACGCGTGGATTCCTCGCTCGCTCGATCCCGCGAGGTGAT 4305  
Qy 205 ---AlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPhe----- 221  
Db 4306 CCGAGCAACCGGTTCGGTGGATCAGCGGCGCACCGCGGACCGCGGCAACCACTATCATCACC 4365  
Qy 222 -----TrpSerArgAspGlyGlyAlaThrTrpGlnAla 232  
Db 4366 GATGGCAACCGCAACCGATTGGCCAGGTTCACCGCGCGCGCGGAGCGGTAACTGGTCTTC 4425  
Qy 233 ValProGlyAlaPro----- 237  
Db 4426 ACTCCAGGCATCCGCTGCGCGGATGGCAACGTTGGTCAACGTTGGTGGCGCGCGCAACG 4485

QY 238 -----ThrGlyPheIleProHisLysGlyVal 246  
Db 4486 AATGTGACAGTCCGCGCGGTGATCACTGTGATGGTGGCCCGCGCGCGGTG 4545  
QY 247 PheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyr--- 265  
Db 4546 ATGATCCGAGCAACCGCACCGAGATAGCGGTACCGCGGAGCGCGCGCGGTGATC 4605  
QY 266 -----AspGlySerSerGlyAspValTrpLysPheSerValThr---SerGlyThrTrp 282  
Db 4606 CTCACCGATGGCGCGCAACCGCATCGGCAGCGCACCGCCAGCGCAACTGG 4665  
QY 283 ThrArgIleSerProValPro-----SerThrAsp 292  
Db 4666 ACCTTACCCCGCGCACCGCGTGGCGCAACCGCATCAACCGCGTGGCCAGGAC 4725  
QY 293 ThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisPro----- 310  
Db 4726 CCGCGCGCAATACCAAGCGGTCCGCGCGCATCGGTCACTGATCCATCGCCCGCGCG 4785  
QY 311 -----AsnThrIleMetValAlaThrGlnIleSerTrpTrpProAsp 324  
Db 4786 CCGGTGATCAATCCGAGCAACCGAGTGGTATCAGCGGTACGCGGAAGCCGGGCGCACG 4845  
QY 325 ThrIleIlePheArgSerThrAspGlyGly-----AlaThrTrpThr 338  
Db 4846 GTGATCCTC-----ACCAGCGCAACCGCATCGGCAGGTCACTCCGCGCAC 4896  
QY 339 ArgIleTrpAspTrpThr-SerTyrProAsnArg-----SerLe 351  
Db 4897 GGCAGCGCAAGTGGGCTTTACGCCCGCCACCGCGTTGGCCAATGGCACCGGTGATCAAT 4956  
QY 351 uArgTyrValLeuAspIleSerAlaGluProTyrLeuThrPheGlyValGlnProAsnPr 371  
Db 4957 GGCTGG-----CCAGAGCGCGCGCGCAACAGCGATC-----CC 4995  
QY 371 oProValProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAs 391  
Db 4996 ACCAGCGCCA---CGGTGACTCGTGGCGCCAGCA-GCCCGGTGATCGATCGAGCA 5051  
QY 391 nSerAspArgMetLeuTyrGlyThr-----GlyAlaThrLeuTyrAlaThrAsnAs 408  
Db 5052 CGGTAGC---GTGATCGCGGTACCGCGAGGTGGTGCACCGGTGATCCTCACC----- 5103  
QY 408 pLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuCl 428  
Db 5104 -----GACGCAACGGCAAC-----CCGATCGCGCAGGTCAACGCG 5138  
QY 428 uGluThrAlaValAsnAspLeuIleSerProSerGlyAlaProLeu----- 444  
Db 5139 CGATGGCAGCGCAACTGGAGCTTCACGCCC-----GGCAGCGCGCTGTCCAATGGCAC 5192  
QY 445 -----IleSerAlaLeuGly---AspLeuGlyGlyPheThrHisAlaAspVal---ThrAl 461  
Db 5193 GGTGTCAATGGGTGGCGCGAGCGCTCCCGCAACACCGCGCGCGCGCGCACCAAC 5252  
QY 461 aValProSerThrIlePheThrSerProValPhe-----ThrThrGlyThrSerVal-- 478  
Db 5253 GGTGACTCGGTGGCGCGCGCGCGGTGATCGACCGCAACCGCATCGGTGATCGC 5312  
QY 479 -AspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerSe 498  
Db 5313 CGGTACCGCGGAAGCGGTGGCGAGGTGATCTCACCAGTGGCGCGCGCAACCCGATCGG 5372  
QY 498 rGlnProAsnAspArgHisValAlaPheSerThrAspGlyLysAsnTrp----- 515  
Db 5373 CCAGGCC-----ACCGCGATGGCAGCGCAACTGGAGCTTCAC 5411  
QY 516 -----PheGlnGlySerGluProGlyGl 523  
Db 5412 CCGCGGCGCGCGGTGGCGCAACCGCACCGGTGATCAATCGGTGGCGCGAGGATCCGCGCG 5471  
QY 523 yValThrThrGlyGlyThr----- 529

Db 5472 CAATACCAAGCGCGCGCACCGGTGGACCGGTGGACCGCGCGCGCGCGCGCGCGGTGGT 5531  
QY 530 -----ValAlaAlaSerAlaAspGlySerArgPheValTrpAl 542  
Db 5532 CAACCCGAGCAACCGCACCGGTGATCGCGGTATCCGCGGAAGCGCGCGCGCGGTGATCT 5591  
QY 542 aProGlyAspProGlyGlnProVal-----ValTyrAlaValGlyPheGlyAsnSerTr 560  
Db 5592 CACGAGCGCGCGCGCAACCGCATCGGCCAGGTCAACCGCGCACCGCGCGCGCAAC--TG 5648  
QY 560 pAlaAlaSerGlnGlyValPro---AlaAsnAlaGlnIleArgSer----- 574  
Db 5649 GAGTTTACCGCGCGCACCGCGTGGCGCAACCGGTTCGTGATCAATGCGTGGCGCGCACGA 5708  
QY 575 -----AspArgValAsnProLysThr 581  
Db 5709 CGCGCGCGCGCAACACCGACCGCGCGCGCACCGGTGACTCGGTAGCCCGCGCGCAC 5768  
QY 581 r---PheTyrAlaLeuSerAsnGlyThrPheTyrArgSer----- 593  
Db 5769 CCGGTGCTCGATCCGAGCAACCGGTAGCTGATCAGCGGTACCGCGAAGCGCGCGCAC 5828  
QY 594 -----ThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeu-ProSers 610  
Db 5829 GGTGATCCTCACCGACCGCGC-----GGCAACCCGATACCGCGCGCACCGCGCGGTG 5882  
QY 610 erGlyAlaValGlyValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuA 630  
Db 5883 CAGCGCAACTGGAGTTTCACTCCGCGGCACA-CGGCTGACCAACCGCACCGGTGATCAATG 5941  
QY 630 lAlaSerSerGlyLeuTyrHisSerThrAsnGly-----GlySerSerTrpSerAlaI 648  
Db 5942 CGGTGCGCGAGCGCGCGCGCAACACCGCGGTCCGCTCAGCACCGACAGTGGACGCGG 6001  
QY 648 leThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerT 668  
Db 6002 TGCGCGCGCGCGCGCGGTGATCGCCGAGCAATGTGTCAAACTCAGCGCGCACCGCG 6061  
QY 668 yProAlaValPheValValGlyThrIleGlyGly----- 679  
Db 6062 AACCCGCGCTCGGGTGTCTCACCAGTGGCAATGGCAACCGCATCGCGCGCACCGCTCG 6121  
QY 680 -----ValT 681  
Db 6122 CCGACGCTAGCGCAACTGGACCTTCACCGCGCGCACCGCGTGGCCAAACCGCGCGTGG 6181  
QY 681 hrGlyAlaTyrArgSerAspAspCysGlyThrThrTrpValLeuIleAsnAspAspGlnH 701  
Db 6182 TCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCAATACCAAG-----CGTCCGCGCGC 6231  
QY 701 ieGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgArg-Val 720  
Db 6232 ACCCGGTGATACGTGGCGCGCGCGCGCGCGCGCGGTGATCAATCCGCAACCGCGCGGTG 6291  
QY 721 TyrIleGlyThrAsn-----GlyArgGlyIleValTyrGlyAspIleGlyAlaPro 738  
Db 6292 ATCAGCGCGCACCGCGAGGTGGCGCGCAAGTGTCTCACCAGCGCGCAACCGCGCACCG 6351  
QY 739 SerGly 740  
Db 6352 ATCGCG 6357

## RESULT 3

US-09-252-991A-13774/c  
; Sequence 13774, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A

Qy	261	ThrGlyGlyProTyr-----	AspGlySerSerGlyAspValTrpIysPheSerVal	277
Db	1821	CGGGCAGCAGCGTGACCTGACCGATGGCAACGCAACCCGATCGCCAGGTCAACGCC	1762	
Qy	278	Thr---SerGlyThrTrpThrArgIleSerProValPro-----	289	
Db	1761	GACGGCAGCGGCAACTGGAGCTTCACCCCGTCACACCCCGTGGCGGATGGAAACCGTGTC	1702	
Qy	290	-----SerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAsp	306	
Db	1701	AACGCCACGCGCACCGATCCGGCGGGCAACACACAGCGGCCAGGGCAGCACCACCGTCGAT	1642	
Qy	307	ArgGlnHisProAsnThrIleMetVal-----	AlaThrGlnIleSerTrpTrp	322
Db	1641	GGCGTGGCGCCGACGACCGCGCTCAACCTGAGCAACGGCAGCAGCCTCAGCGGCAC	1582	
Qy	323	ProAsp-----ThrIleIlePheArg-----	329	
Db	1581	CGCGAACCAGCGCAGCAGCGGTGATCTCACCGACGGCAACGGCAATCCGATCGCCGAGGTC	1522	
Qy	330	SerThrAspGlyGlyAlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArg	349	
Db	1521	ACCGCCGACGCGCAGCGCACTGGACC-----TACACCCCGT--CCACGCCGA	1476	
Qy	350	SerLeuArgTyrValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnPro	369	
Db	1475	TCG-----CCAACGGCACCGTGGTCAACGTGGTGGCCAGGACG	1437	
Qy	370	AsnProProValPro-----	SerProLysLeuGlyTrpMetAspGluAlaMetAla---	386
Db	1436	CCGCGCGCAATAGCAGCGCGGCGCCAGCGTCAACCGTGGAC--TCGCAAGCCCGGGGGCT	1378	
Qy	387	-----IleAspProPheAsnSerAspArgMetLeuTyrGlyThr-----	GlyAla	401
Db	1377	CCGGTGGTCAACCCGAGCAACGCG--ACCACGCTCAGCGGCACCGCCGAGCGGGCGCT	1321	
Qy	402	ThrLeuTyrAlaThrAsn-----	AspLeuThrIysTrpAspSer	414
Db	1320	ACCGTGAGCTGACCGACGGCAACCGCATGGCCAGGTCAACGCC--GACGGC	1264	
Qy	415	GlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluThrAlaValAsnAsp	434	
Db	1263	AGTGGCAACTGGAGCTTACACCGGGCAGCGCCGTGGCCAAACGGCACCGTGGTCAACGCC	1204	
Qy	435	LeuIleSerProProSerGly-----	AlaProLeuIleSerAlaLeuGlyAspLeu	451
Db	1203	ACGGCAGCAGCCGACCGGCAATACAGCGCTCCCGGCAGCACCCGTTGGATCGGTG	1144	
Qy	452	GlyGlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProVal	471	
Db	1143	-----GGCGCGCGCGCCGCTGGTGGTCAATCGGACCAACGAGTC	1105	
Qy	472	PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAla	491	
Db	1104	GTCAATCAGCGGCACCC-----GCCGAACCGGGGCCACCGTGACCTGACCGAT	1057	
Qy	492	GlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly	511	
Db	1056	GGCAGCGGCAATCCGATCGGGCAG-----GTCACCCCGCGCAGCC	1018	
Qy	512	GlyLysAsnTrp-----	Phe	516
Db	1017	ACGGCAACTGGAGCTTACCCCGTCCACGCCCGTGGCGGATGGAAACCGTGGTCAACGCC	958	
Qy	517	GlnGlySerGluProGlyGlyValThrThrGlyGly-----	ThrValAlaAla	532
Db	957	ACCGCTACCGACCCGCGCGGC--AATACCGCGGCCAGGGCAGCAGCTACCGTGGACGCC	901	
Qy	533	SerAla-----	AspGlySerArgPheValTrpAlaPro	543
Db	900	ATCGCGCGCGCCACGCGCAGCCGTCAACCTGAGCAATGGCAGCAGC--CTCAGCGGCAC	844	

```
QY 544 GlyAspProGlyClnProValValTyrAlaValGlyPheGlyAsn----- 558
Db 843 GCGGAACCGCGGACGACGGTGTCTCACCAGCGCAACGCAATCCGATCGCGGAGTGC 784
QY 559 -----SerTrp----- 560
Db 783 ACCGCCGACGCGACGCGCAACTGACCTACCCCTCCACGCCGATCGCAACGGTACT 724
QY 561 -----AlaAlaSerGlnGlyValProAlaAsnAlaGln 571
Db 723 GTGGTCAACGTGTGGCCGAGGACCGCGGTAAACAGCAGCCCGCGCGAGGTGACC 664
QY 572 IleArgSerAspArg-----ValAsnProLysThrPheTyrAlaLeuSer 586
Db 663 GTCGATTTCAGCGCGCGCGCGCGCGGTGATCAACCCG-----AGC 622
QY 587 AsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProValAlaAlaGly 606
Db 621 ACGCGCGTC-----GTCATCAGCGGC 601
QY 607 LeuProSerSerGlyAlaValGlyValMetPheHisAlaValProGlyLysGluGlyAsp 626
Db 600 ACGCGCGAGCGCGTCCACCGTACCCCTCACCAGTCCCGCGCAACCCGATAGGCGAG 541
QY 627 LeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSer 646
Db 540 GTC-----ACGCGCGACGCGACGCGCAACTGGAGC 511
QY 647 AlaIleThrGlyValSerSerAla-----ValAsnValGlyPheGlyLysSerAla 663
Db 510 TTCACGCGCGGACGCGCGCGGCAACGCGACGGTGTCTGCCACCGCCACCGACCCG 451
QY 664 ProGlySerSerTyrProAlaValPheValValGlyThrIleGlyGlyValThrGlyAla 683
Db 450 ACGGCAATACCGCGCGCGCGCGCGCCACAC-GGTGGACGGTGGCGCGCGCGCGCC 392
QY 684 TyrArgSerAspCysGlyThrThrTrpValLeuIleAsnAspAspGlnHisGlnTyr 703
Db 391 -----GGTGTATCGATCCGACCAACGCGACCATCAGCGGCGCGCC 350
QY 704 GlyAsnTrpGlyClnAlaIleThrGlyAsp-----HisAlaAsnLeuArgArg 719
Db 349 GGAGCGCGCGGCGCAA-----GGTGTCTCTCACCAGCGCAACGCGCAA 308

RESULT 4
US-09-252-991A-13773/c
; Sequence 13773, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13773
; LENGTH: 4131
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13773

Alignment Scores:
Pred. No.: 2,92e-08 Length: 4131
Score: 201.50 Matches: 181
Percent Similarity: 31.62% Conservative: 78
Best Local Similarity: 22.10% Mismatches: 274
Query Match: 4.99% Indels: 287
pb: 4 Gaps: 43
```

```
US-09-917-376-3 (1-740) x US-09-252-991A-13773 (1-4131)
QY 48 AsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr----- 65
Db 3954 AACGGCAACCGGATGGCCAGGTCCCGCGACGCGGTAAGTCTTCTTCACTCCA 3895
QY 66 -----AsnGlyValVal-----SerIleAlaAlaAspProIleAsnThr 78
Db 3894 GGCATCCCGCTCGCGGATGGCACGGTGGTCAACGTGGTGGCGCGCAAGCAATGTC 3835
QY 79 AsnLysValTrpAla-----AlaValGlyMetTyrThrAsnSerTrpAsp 93
Db 3834 GACAGTCCGCGCGCGGTGATCACTGTGGTGGCGCGCGCGCGCGCGGTGATCGAT 3775
QY 94 ProAsnAspGlyAla-----IleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThr 112
Db 3774 CCAGGACACCGCACCGAGATAAGCGGTACCGGAGCGCGCGCGCGGTGATCTCTACC 3715
QY 113 ProLeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAla 132
Db 3714 GAT-----GGCGGGCGCAACCG-----ATCGGCAGCGCCACCGCC 3679
QY 133 ValAspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrp 152
Db 3678 -----GACGCGCAGCGCGCAACTGG 3661
QY 153 ArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThr 172
Db 3660 ACCTTACCCCG-----GGCACC 3643
QY 173 TyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrp 192
Db 3642 CCGCTGGCAACCGCACCGTATCAACGCG----- 3613
QY 193 ValAlaPheAspLysSerSerSerSerLeuGlyGlnAlaSerLysThrIlePhe----- 210
Db 3612 GTGGCCAGGACCGCGCGCGCAATACCGCGGTCCGCGCGCGCGCGCGTCACTCGTGCATC 3553
QY 211 -----ValGlyValAlaAspProAsnAsnProValPheTrpSer----- 223
Db 3552 GCCCGCGCGCGCGGTGATCAATCCAGCAACGAGTGTGTATCATCAGCGGTACGCGGAA 3493
QY 224 -----Arg 224
Db 3492 GCCGGGCGCACGGTGTATCTCTACCGACGCGCAACGCGATCGCGCAGGTACCGCC 3433
QY 225 AspGlyGlyAlaThrTrpGlnAlaValProGlyAlaPro----- 237
Db 3432 GACGGCAGCGCAAGTGGGCTTTCACGCGCGCACGCGGTGGGCAATGGCAGCGTATC 3373
QY 238 -----Thr 238
Db 3372 AATGCGTGGCCCGACGACGCGCGCGCAACACAGCAGTCCCACAGCGCCACCGTGCAC 3313
QY 239 GlyPheIleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThr 258
Db 3312 TCCTGGCGCAGCAGCGCGCGGTGATCGATCCGAGCAACGGTAGCGGTGATCGCGGTACC 3253
QY 259 SerAsnThrGlyGlyProTyr-----AspGlySerSerGlyAspValTrpLysPhe 275
Db 3252 GCCGAGGTGTGTCACGGTGTCTCACCAGCGCAACGCAACCGCATCGCGCAGGTGC 3193
QY 276 SerValThr-----SerGlyThrTrpThrArgIleSerProValPro----- 289
Db 3192 ACCGCCGATGGCAGCGCAACTGGAGCTTCACGCCCGCGCGCGGTGTCTCAATGGCAGC 3133
QY 290 -----SerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThr 304
Db 3132 GTGGTCAATCGGTGGCGCGCGGTGCGCGCAACACGACGCGCGCGCGCGCAGCACCG 3073
QY 305 IleAspArgGlnHisPro-----AsnThrIleMetValAla 316
Db 305 -----AsnThrIleMetValAla 316
```

Db 3072 GTGACTCGTGGCGCGCGCGCGGTGATCGACCCGAGCAACGGCGGTGATCGCC 3013  
Qy 317 ThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGly 334  
Db 3012 GGTACCGGGAAGCGGTGCGACGGTATCTC-----ACGATGGCGGCGCAAC 2962  
Qy 335 -----AlaThrTrpThrArgIleTrpAspTrpThrSerTrpProAsnArgSer 350  
Db 2961 CCGATCGCGCCAGGCGCGCGATGGCAGCGCAACTGGAGTTTCACCCCGGCGACGCG 2902  
Qy 351 LeuArgTrpValLeuAspIleSerAla-----Glu 360  
Db 2901 CTGGCCAAAGCGACGGTATCAATCGGTGGTGGCCAGAGTCCGGCGGCAATACACAGCGCG 2842  
Qy 361 ProTrpLeuThrPhe-----GlyValGlnProAsnProProValProSerProIlysLeu 378  
Db 2841 CCGACCAACCGACGGTGGAGCGGTGGCCCGCGCCACCGCGGTGGTCAACCCGAGCAAC 2782  
Qy 379 GlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTrpGly 398  
Db 2781 GGC-----AGCGTGTATCGCGGT 2764  
Qy 399 Thr-----GlyAlaThrLeuTrpAlaThrAsnAspLeuThrLysTrpAspSerGly 415  
Db 2763 ACCCGGAAGCGCGCGCGCGGTATCTCACC-----GACGGCGCG 2722  
Qy 416 GlyGlnIleHisIleAlaProMetValLysGlyLeuGluGluThrAlaValAsnAspLeu 435  
Db 2721 GGCAC-----CCGATCGCGCGGTACCGCGCGAGCGCGGCAACTGGAGC 2674  
Qy 436 IleSerProProSerGlyAlaProLeu-----IleSerAlaLeuGly 449  
Db 2673 TTCACGCCC-----GGCACCGCGTGGCCAAACGGCTCGGTGATCAATGCGTGGCCCGAG 2620  
Qy 450 AspLeuGlyGlyPheThrHisAlaAspVal-----ThrAlaValProSerThrIlePheThr 468  
Db 2619 GACGCGCGCGCAACACACGAGCGCGCGCGCGCGCGGTGAGTCTCGGTAGCCCGCGCC 2560  
Qy 469 SerProValPhe-----ThrThrGlyThrSerValAsp-----TyrAlaGluLeuAsnPro 485  
Db 2559 ACCCGGTGCTCGATCGGCAACGGTACGGTATCGAGCGGTACCGCGCAAGCGCGCGCC 2500  
Qy 486 SerIleIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArgHisVal 505  
Db 2499 ACGGTGATCTCCAGCGCGCGCGCGCGCAACCGATACCGCAGCGCC-----2455  
Qy 506 AlaPheSerThrAspGlyGlyAsnTrpPheGlnGlySerGluProGlyGlyValThr 525  
Db 2454 -----ACCGCGGATGGCAGCGGCAACTGG-----AGTTCACTCCGGGCGACACCGCTG 2407  
Qy 526 ThrGlyGlyThrVal-----AlaIleSerAlaAspGlySerArgPheValTrpAlaPro 543  
Db 2406 ACCAAACGCGCGGTGATCAATGCGGTGGCCCGCGCGCGCGCGCAACACCGCGGTCCG 2347  
Qy 543 -----543  
Db 2346 GTCACACACAGTGGAGCGGTGGCCCGCGCCACCGCGGTGATCGACCCGAGCAATGGT 2287  
Qy 544 -----GlyAspProGlyGlnProValValValTrpAlaValGlyPheGly 557  
Db 2286 GTCAAACTCAGCGGACCGCGCAACCGCGGTGCGGGTGTCTCACCAGTGGCAATGGC 2227  
Qy 558 Asn-----SerTrpAlaIleSerGlnGlyVal 566  
Db 2226 AACCCGATCGGCCAGACCTCGCGACGGTACGCGCAACTGGACCTTCACACCGGGGACG 2167  
Qy 567 ProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrAlaLeuSer 586  
Db 2166 CCG-----CTGGCC 2158  
Qy 587 AsnGlyThrPheTyrArg-----SerThrAspGly-----596  
Db 2157 AACGGCAGGTGGTCAACCGCGGTGGCCCGCGGAGCGCGCGGCAATACCGCGGTCCGCGCC 2098

Qy 597 GlyValThrPheGlnProValAlaAlaGlyLeuPro-----SerSerGlyAla 612  
Db 2097 AGCACACCGGTGATACGGTGGCGCGCGCGGTGATCAATGATCGACCAACGCGGACG 2038  
Qy 613 ValGlyValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaIleSer 632  
Db 2037 GTG-----ATCACCGCGCACCGCGCGGTGGCGGCAAGTATCTCACCGCGGCAACGCG 1981  
Qy 633 SerGlyLeuTyrHisSerThr-----AsnGlyGlySerSerTrpSerAlaIleThrGlyVal 651  
Db 1980 AACCCGATCGCGGAGACACCGCGCGCGAGTGGCACTGGACCTTCACCCCGCGGACG 1921  
Qy 652 -----SerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSer 667  
Db 1920 CGCGTGGCCAAACGGTACGGTATCAACGGCGTCCGCGAAGACCGCGGCGCAACGCGCAG 1861  
Qy 668 TyrProAlaValPheValGlyThrIleGlyGlyValThrGlyAlaTyrArgSer 686  
Db 1860 GGTCCGCGCC-----AGCACAC-GGTGGACTCGGTGGCGCGCTCCCGCTCC 1817

RESULT 5  
US-09-103-840A-1/c  
; Sequence 1, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 4411529  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; OTHER INFORMATION: H37Rv  
US-09-103-840A-1

Alignment Scores:  
Pred. No.: 0.0104 Length: 4411529  
Score: 192.00 Matches: 194  
Percent Similarity: 35.04% Conservative: 94  
Best Local Similarity: 23.60% Mismatches: 323  
Query Match: 4.76% Indels: 213  
DB: 3 Gaps: 43

US-09-917-376-3 (1-740) x US-09-103-840A-1 (1-4411529)

Qy 10 AsnValAlaIleGlyGlyGlyGlyPheValAspGlyIleValPheAsnGlyGly-----27  
Db 427279 AACGGAACATCGCGCGCGCAACATCGCGAC-----TTTAACGTGGATCGCA 427229  
Qy 28 -----AlaProGlyIleLeuTyrValArgThrAspIleGly-----GlyMet 41  
Db 427228 AACACCGGTCCGGGCTAACGCGGCTGTCAACACATCGGTATCGGCAACACCGGCAAC 427169  
Qy 42 TyrArgTrpAspAlaAlaAsn-----GlyArgTrpIleProLeuLeuAspTrpValGlyTrp 60  
Db 427168 TACAACATCGGTTCGGCAACACCGGTAACTACAAC-----ATCGGTTC 427124  
Qy 61 AsnAsnTrpGlyTyrAsnGlyVal-----ValSerIleAlaAlaAsp-----74  
Db 427123 GGCACACCGCGCAACACACATCGCATCGGCTGTCCGGCGCAACACCATCGGTTC 427064  
Qy 75 ---ProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAsp 93  
Db 427063 GGCCCGCTGAACCGCGCC-----ATCGCAACATCGGCTGTTC---AACCTGGCGAC 427013

```
QY 94 ProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrPro 113
Db 427012 AACAACTTTGGC----- 427001

QY 114 LeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaVal 133
Db 427000 -----ATGGCCAAACGGGGCAACTTCAACACCGGGCAATTGCC 426965

QY 134 AspProAsnAsnAspAsnIleLeuTyPheGlyAlaProSerGlyLys---GlyLeuTrp 152
Db 426964 AACCGGCAACAACAACATCGCTGTTCAACACCGGCAACAACAACATCGGCATCTCG 426905

QY 153 ArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThr 172
Db 426904 CTGACCGGCGAGCGGTGTCCTCGGCTTCAGCTCCCTGAACCTCGGCGCGGCAACACCGGT 426845

QY 173 TyrlleAlaAsnProThrAspThrThrGlyTyGlnSerAspIleGlnGlyValValTrp 192
Db 426844 TTCTTCAACTCCGGCACCGCCCAACACCGGC----- 426815

QY 193 ValAlaPheAspLysSerSerSerSerLeuGlyGln---AlaSerLysThrIlePheVal 211
Db 426814 ---TTGTTCAACTCCGGCACCGCAACACCGGCTTGTTCAACTCGGGCACCGCAACGTC 426758

QY 212 GlyValAlaAspProAsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGln 231
Db 426757 GGCATCGGCAACATGGGCACCGCGGCTTCGGCGTTCGGCTATCCGGCGACAGCCAGGTG 426698

QY 232 AlaValProGlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsn 251
Db 426697 GGCATCGGCGGCAACCAACTCGGGCAGTTTC---AACTCGGCTGTGTTAACTCGGGCAC 426641

QY 252 HisValLeuTyrlleAlaThrSerAsnThrGlyGlyProTyAspGlySerGly-As 271
Db 426640 GGCATGTGCGCATCGGCAACTCGGGCACCGGCAAGCTCGGCATCGGCAACCGGCACC 426581

QY 271 pValTrpLysPheSerValThrSerGlyThr-----Tr 282
Db 426580 GGCAACACCGGCATCGGAACACGGCGCAACTACAACACCGGCTGTGTCACCGGGGCGCTG 426521

QY 282 p-----ThrArgIleSerProValProSerThrAspThrAlaAsnAs 296
Db 426520 GTCAACACCGGCATCGCCAAACCGGCGCAACCAACACCGGCTGTGTCACCACTCGGCACC 426461

QY 296 pTyPheGlyTySerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAl 316
Db 426460 TTCAACACCGGCATCGCCAAACCGGCGCAACTACAACACCGGCTCTTCAACACCGGTAGC 426401

QY 316 aThrGlnIleSerTrp---TrpProAspThrIleIle---PheArgSerThr----- 331
Db 426400 TACAACACCGGCATCGCAAAACCGGCGAGACTAGCGCACCGCGCGTTCATACCGGCGCAGC 426341

QY 332 -----AspGlyGlyAlaThrTrpThrArgIleTrpAspTrpThrSer--- 345
Db 426340 ATGAACAACCGGCTGTCTCGGCGCGCGGCGGCGGCGGCTGTCTGCGGCGCAACTACCC 426281

QY 346 -----TyPProAsnArgSerLeuArgTyValLeuAspIleSerAlaGl 360
Db 426280 ATCACCATCGACGGACCTCGCGGTTCCTCAATGTCGACATCCCGGTCAACATCCCATC 426221

QY 360 uProTrpLeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGlyTr 380
Db 426220 ACCG---GGACATCACCAATGTCTCCATCCCGCCATACGTTCCCCAGA-ATC----- 426170

QY 380 pMetAspGluAlaMetAlaIleAsp---ProPheAsnSerAspArgMetLeuTyGlyTh 399
Db 426169 ----GACGCGCAGGAGCGGTGACATAGCATGTCCTCACTGGCAGCGCTTGGCCCGCGGT 426114

QY 399 rGly---AlaThrLeuTyAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnI 418
Db 426113 CGGTCCGATCACCCCTGCATGGGGGACGCGGTGCGGCCCGCGCTGGACACACCCATCGAAAT 426054
```

RESULT 6



```

US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R. M.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Alignment Scores:
Pred. No.: 0.0564 Length: 4403765
Score: 184.00 Matches: 204
Percent Similarity: 32.65% Conservative: 116
Best Local Similarity: 20.82% Mismatches: 306
Query Match: 4.56% Indels: 361
DB: Gaps: 50

US-09-917-376-3 (1-740) x US-09-103-840A-2 (1-4403765)

Qy 14 GlyGlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeu--- 32
Db 3745039 GCGGTGGCGGTCAACCGGATGCG-----GATCGCGCGGTGGCGGTGCGAATCT 3744989
Qy 33 -----TyrValArgThrAspIleGly 39
Db 3744988 GTTGGGCGAGAACGCCGACAGATCGCGCGCTCGAGCGCGAGTACGATGATGGCG 3744929
Qy 39 ----- 39
Db 3744928 CGCGATGTGGCGGCGATGCGCGCTACCATTCGGCGCGGTCTGCTCCCGGGCGGT 3744869
Qy 40 -----GlyMetTyrArgTTPAspAlaAlaAsnGlyArg----- 50
Db 3744868 GCGGCGGTTTCAGCCACCGCGCGCAGG-CGTGGGGGTGGTGTGCGCGCGTCTCTCAATG 3744810
Qy 51 -----TTP-----TTP-----IleProLeu 54
Db 3744809 CTCATTTCGCGACCGCGAGATGTTAGGCTTAACGCGCGCTTGGGCAATGTCGGTA 3744750
Qy 55 LeuAspTTPValGlyTTPAsnAsnTTPGlyTTPAsnGlyValValSerIleAlaAlaAsp 74
Db 3744749 ATTACACGTCGGTGGGCAATGTCGGATAT-----TCAACCTGGGCGCAG 3744702
Qy 75 ProIleAsnThrAsnLysValTTP-----AlaAlaValGlyMetTyr 88
Db 3744701 CCAATGTCGGTCGCGAGAAATTGGGTGCTGCCAACGCGGTAGCGGAATTCGGTTTCG 3744642
Qy 89 ThrAsnSerTTPAspProAsnAspGlyAlaIleLeuArg----- 101
Db 3744641 GCAATATCGCAACGCGCAACTTCGGGTTCGGCAACTTCGGGTCTTGGGTTCGCCCGCGGCA 3744582
Qy 102 -----SerSerAspGlnGlyAlaThr 108
Db 3744581 TGGCAATATTGGTGGGCAATTCGGGCGAGCAGCAACTACGGCTTCGAAACCTGGTG 3744522
Qy 109 TTPGlnIleThrProLeuProPheLysLeu-----Gly 119

```





QY 694 ValLeuile-----AsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThr 711  
 : : : : :  
 Db 4057 ACAGTGTAGCAGCCAGATGTAGTGTAGTACCAACCCAGTAACCTGGTAGAT 4116  
 : : : : :  
 QY 712 GlyAspHisAlaAsnLeuArgValTyrIleGlyThrAsn-----GlyArg 727  
 : : : : :  
 Db 4117 GGTGATACAGTAAGTCAACAGCACTGACCTGCGAGGCAATACATTCATTCGCGAGGTACA 4176  
 : : : : :  
 QY 728 GlyIleValTyrGlyAspIle 734  
 : : : : :  
 Db 4177 GGCACAGTCTCTGCAGACATC 4197  
 : : : : :

RESULT 8

US-09-453-702B-39  
 ; Sequence 39, Application US/09453702B  
 ; Patent No. 6365723  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Blattner, Frederick R.  
 ; Burland, Valerie  
 ; Perna, Nicole T.  
 ; Plunkett, Guy  
 ; Welch, Rod  
 ; TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157  
 ; NUMBER OF SEQUENCES: 265  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Quarles & Brady  
 ; STREET: 1 South Pinckney Street  
 ; CITY: Madison  
 ; STATE: WI  
 ; COUNTRY: US  
 ; ZIP: 53701-2113  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Word Perfect 8.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/453,702B  
 ; FILING DATE: 03-Dec-1999  
 ; CLASSIFICATION: <Unknown>  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 60/110,955  
 ; FILING DATE: 04-DEC-1998  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Seay, Nicholas J.  
 ; REGISTRATION NUMBER: 27386  
 ; REFERENCE/DOCKET NUMBER: 960296.95017  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (608) 251-5000  
 ; TELEFAX: (608) 251-9166  
 ; INFORMATION FOR SEQ ID NO: 39:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 25165  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 39:  
 US-09-453-702B-39  
 Alignment Scores:  
 Pred. No.: 0.000337 Length: 25165  
 Score: 170.50 Matches: 192  
 Percent Similarity: 29.22% Conservative: 83  
 Best Local Similarity: 20.40% Mismatches: 269  
 Query Match: 4.22% Indels: 397  
 DB: 4 Gaps: 48  
 US-09-917-376-3 (1-740) x US-09-453-702B-39 (1-25165)  
 QY 27 GlyAlaProGlyIleLeuTyrValArgThrAspIleGlyMetTyrArgTrpAspAla 46  
 : : : : :  
 : : : : :

Db 18607 GGCCTTAATGAGTTGCGCAAAATCAGTACCGACAAACCGCGCA----- 18648  
 QY 47 AlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAsn 66  
 : : : : :  
 Db 18649 -----ACCTGGGTGAAC----- 18660  
 : : : : :  
 QY 67 GlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaValGly 86  
 : : : : :  
 Db 18661 -----GTGACCGTGGCGCAGACAGCTGAAC----- 18687  
 : : : : :  
 QY 87 MetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGlnGly 106  
 : : : : :  
 Db 18688 -----TGGAGTTAGTTGACGAGCAACCCCTCACACAGGGCACC----- 18726  
 : : : : :  
 QY 107 AlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArgGly 126  
 : : : : :  
 Db 18727 ACCACCTGGCAGGTGGCGGTGTC-----GATCTGGCGGCAAGTT--GGCGCAACG 18777  
 : : : : :  
 QY 127 MetGlyGluArgLeuAlaValAspProAsnAsn----- 137  
 : : : : :  
 Db 18778 AGCAGCCAGTGGCGCTGATCGATACCGTTAAACCGCGCAGGTGCTCACCATCGCCAGC 18837  
 : : : : :  
 QY 138 -----AspAsnIleLeuTyr 142  
 : : : : :  
 Db 18838 ATCAGCACCGACACACGGGAGTTGCGCAACTGACTTTATCACGACGACACCATGCTCAG 18897  
 : : : : :  
 QY 143 Phe-----GlyAlaProSerGlyLysGlyLeuTrpArgSerThrAsp 156  
 : : : : :  
 Db 18898 CTGACCGGTTGCTGGCGGGCGGCTTGCACGCGCGAAGTGGCGCAGATTAGCCTTGAT 18957  
 : : : : :  
 QY 157 SerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyTyrTyrIleAlaAsn 176  
 : : : : :  
 Db 18958 AGCGGCGCGACCTGGACACGCTCACCACCAACGGTACACAGTGGAGCTTACACCGACAGC 19017  
 : : : : :  
 QY 177 ProThrAspThrThrGly-----TyrGlnSerAspIleGlnGlyValValTrpVal 193  
 : : : : :  
 Db 19018 CGCAGCGTACCGACGCGAGCTAGTTTATCAGGTGCGGGTG----- 19059  
 : : : : :  
 QY 194 AlaPheAspLysSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyVal 213  
 : : : : :  
 Db 19060 ---CTGATCTGGCGGGGAACACCGCGCGGTGTGCGAAGCGGTGGTGTGATGATG 19116  
 : : : : :  
 QY 214 AlaAspProAsn-----AsnProValPheTrpSerArgAspGlyGlyAlaThrTrp 230  
 : : : : :  
 Db 19117 ATTAACCCCGCCGACACCAACGATTGTGTCGATATCCGATGATGTCGGG----- 19167  
 : : : : :  
 QY 231 GlnAlaValProGlyAlaProThrGlyPheIleProHisLysGlyValPheAspProVal 250  
 : : : : :  
 Db 19168 -----CAGCGCGCAGGGGACATTAAAGCAGTTTG 19194  
 : : : : :  
 QY 251 AsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAspGly----- 267  
 : : : : :  
 Db 19195 CAG-----GCCACCGACGACACTACCGCGGTGTCGACGGGTGACTTTC 19239  
 : : : : :  
 QY 268 -----SerSerGlyAspVal----- 272  
 : : : : :  
 Db 19240 GCGCGGCTTGCACGCGTGAAGTGTTCCTCTACCGTAACGGCGTCTGTGTTAGGGCG 19299  
 : : : : :  
 QY 273 -----TrpLysPheSer-----ValThrSerGlyThr 281  
 : : : : :  
 Db 19300 GTGACGATGGTGGCGGCTCTGAACCTGACGACGACGCGGGGTGGTGGCGGTGCC 19359  
 : : : : :  
 QY 282 TrpThr-----ArgIleSerProValProSerThrAspThrAlaAsnAspTyrPhe 298  
 : : : : :  
 Db 19360 TATACCTACAGCGCGGAGTGGTGGATTTGGCGGGGAATATCATCTCTCCAGTGATTTT 19419  
 : : : : :  
 QY 299 GlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThr--- 317  
 : : : : :  
 Db 19420 -----GTCCTGACGGTGCATACCTCTATTCCGACACGCTGGCGCAATCACCAGC 19470  
 : : : : :  
 QY 317 ----- 317  
 : : : : :  
 Db 19471 CAGACCGCGCGGATACGCGCGATTATTAGCGGGGTGATCAGCGCGGCTTCCAGT 19530  
 : : : : :

```
QY 317 ----- 317
Db 19531 GGGCAGTATGTTGAAGTGGTATCAACGGCAAAACCTACACCTCTGAACCGGGCGCGCG 19590
QY 318 -----ProAspThr----- 325
Db 19591 GTAGTGGTTCGATCCGGCGCACAAACCTGGTATGTACAGTTCCGGATACCGATCGGCTG 19650
QY 325 ----- 325
Db 19651 ACAGTTTCGGCGCGCCCTATACCGTTACTGCGCAGGTAAAGTTCCGGCGGTAAACGGC 19710
QY 326 -----IleIlePheArgSerThrAspGlyGlyAla 335
Db 19711 AATAAGGCCAATATTAGCAACGGCAGGTGACGGTTAAACGGCGGATTTACACACCG 19770
QY 336 ThrTrpThrArgIleTrpAspTrpThrSerTyProAsnArgSerLeuArgTyValLeu 355
Db 19771 ACCTGGACTACCGCCAGCAAAACACCGCCCTGG-----GGGCTGACCTACCGCCCTC 19821
QY 356 AspileSerAlaGluProCtyrLeuThrPheGlyValGlnProAsnProValProSer 375
Db 19822 GAC---TGCGACGGATGTGGACGGTGTGGCAACACGACGATATGCAATCGACTGAC 19878
QY 376 ProLysLeuGlyTyrMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMet 395
Db 19879 CCA---CTCACTGTCTGAAGACCGCGCTGAC----- 19908
QY 396 LeuTyrglyThrGlyAlaThrLeuTyAlaThrAsnAspLeuThrLysTrpAspSerGly 415
Db 19909 CTGTATCAGCGCGCAACAAC---TACGCCACAGCTCCATTGCCGATTACGACCGTAAC 19965
QY 416 GlyGlnIleHisIleAlaProMetValLysGlyLeuGluGluThrAlaValAsnAspLeu 435
Db 19966 GGC----- 19968
QY 436 IleSerProSerGlyAlaProLeuIleSer-----AlaLeuGlyAspLeu 451
Db 19969 -----ACGGCGATCTGTTTATCACCCTGATGATCAGGTACGGGCTATATT 20016
QY 452 GlyGlyPheThr---HisAlaAspValThrAlaValProSerThrIlePheThrSerPro 470
Db 20017 AACGGCTTTACCAATAACCGCGATGCGACC-----TTTCCAGCGCT 20058
QY 471 Val---PheThrThrGlyThrSerValAspTyAlaGluLeuAsnProSerIleleVal 489
Db 20059 ATTCAGGTCAACGTCGCGCACCTGACGTGGTACGGC-----TCGATTGTG--- 20103
QY 490 ArgAlaGlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThr 509
Db 20104 -----GCATTGTATTAAGAGGGC----- 20121
QY 510 AspGlyGlyLysAsnTrpPheGlnGlySerGluProGly----- 522
Db 20122 GACGGCTATCTCGACTCTCGATTGTTGATGCGTGGCGCGCGGACTCCAAACACCTTCCTG 20181
QY 523 -----GlyValThrThrGlyThrValAlaAlaSerAlaAspGlySerArg 538
Db 20182 TGGAAACACGAGCGACGCTGGTAGGCACTCCACCACGTCGAACACGCGCGGTAGCGCC 20241
QY 539 PheValTrpAlaProGlyAspProGlyGlnProValValThrAlaValGlyPheGlyAsn 558
Db 20242 ACGGTGGCGCGG-----GCGGTGACGGGGTATCTT 20271
QY 559 SerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgVal--- 577
Db 20272 TGCTCAACGAGGTCTGGCGCTCGATCTGAACATGACGCGGAGGATCGACCTGGTTGAG 20331
QY 578 -----AsnProLysThrPheTyAlaLeuSer-----AsnGly 588
Db 20332 CACACCTATAACCTGAACAACCTATTACAGCTGTCTTCGCTCATCAACCGGGGAATGGG 20391
```

Alignment Scores:

Pred. No.:

1.09

Length:

4403765

Score:

170.00

Matches:

171

## RESULT 9

US-09-103-840A-2

Sequence 2, Application US/09103840A

Patent No. 6294328

GENERAL INFORMATION:

APPLICANT: FLEISCHMAN, Robert D.

APPLICANT: WHITE, Owen R.

APPLICANT: FRASER, Claire M.

APPLICANT: VENTER, John C.

TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

TITLE OF INVENTION: TUBERCULOSIS

FILE REFERENCE: 24366-20007.00

CURRENT APPLICATION NUMBER: US/09/103,840A

CURRENT FILING DATE: 1998-06-24

NUMBER OF SEQ ID NOS: 2

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 2

LENGTH: 4403765

TYPE: DNA

ORGANISM: Mycobacterium tuberculosis

FEATURE:

OTHER INFORMATION: CDC 1551

OTHER INFORMATION: "n" bases at various positions throughout the sequence

OTHER INFORMATION: represent a, t, c or g

US-09-103-840A-2

Percent Similarity:	33.42%	Conservative:	82
Best Local Similarity:	22.59%	Mismatches:	324
Query Match:	4.21%	Indels:	184
DB:	3	Gaps:	32
US-09-917-376-3 (1-740) x US-09-103-840A-2 (1-4403765)			
QY	82	TpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArg	101
DB	3791642	TGGCGCGCGCAAGCCTGGCTACGCCAGGCGCATCAACCATCT	CGA 3791686
QY	102	SerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsn	121
DB	3791687	GAAGCTGATCGG	ACGCACCGAGCGGTGGTGA 3791719
QY	122	MetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsn	137
DB	3791720	CCAGGCGGTAGCGCGCTGGCCGGGT	AATGCCGAGCTCAAGCAGCGCTGTGTGTCAG 3791776
QY	138	AspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAspSer	157
DB	3791777	TC-TCCACGATCGCAGCTCGCCGATCGCAAGGAGCAGATTGGCAAGCGCGTGTGAGT	3791835
QY	158	GlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnPro	177
DB	3791836	TCGGCTACAAAGGCCAGGTGCTGCACAAACCGCGGTGTATCTCTGG	3791883
QY	178	ThrAspThrThrGlyTyrGlnSerAspIleGlnGlyVal	ValTrpValAlaPhe 195
DB	3791884	ACACAGCTGTGAGCTCGGAACCCCGCAGATGACCGCAATTGGCACCCGCCA	3791937
QY	196	AspLysSerSerSerLeuGlyGlnAla	SerLysThrIlePheValGlyVal 213
DB	3791938	TCGAACGGATCAGCGCGCACCGCGCGCCACCGCGGAGTACCGCTGTATCGGGCT	3791997
QY	214	AlaAspProAsnAsnProValPheTrpSerArgAspGlyGly	AlaThrTrp 230
DB	3791998	CGGGAGACGATCGGTGCAAGATGATCTCCACAGCTCGGGGTGCGCAACGTGGCCATCC	3792057
QY	231	GlnAlaValProGlyAlaProThrGlyPheIleProHisLysGlyValPheAsp	248
DB	3792058	CAGCGAAGAGCAACCCAGCGCCACCGCGCGCGCATTCGNACACCGACGGG	3792108
QY	249	ProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySer	268
DB	3792109	CATTCCGGGACAAAGATCAATGGCGAACCG	GATCCGAAG 3792147
QY	269	SerGlyAspValTrpLysPheSerValThrSerGlyThrTrpThrArgIle	285
DB	3792148	GACGCATCAACCACTCAAGCGCAGCTACGGCTGGAACCGCACCGCATCCCGGCATCA	3792207
QY	286	SerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThr	304
DB	3792208	CCGGCCCGCAACCTGGTGGAGACCGGGCTCT	TCGCCCAACAC 3792252
QY	305	IleAspArg-GlnHisProAsnThrIleMet	ValAlaThr 317
DB	3792253	TCGTCAAGATCAGACCCCTGGCAGGTGACAGACACCCCGCCACCCCGCACCGCCAC	3792312
QY	317	glnIleSerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThr	336
DB	3792313	GCAGGTGCGC	CAGCGCGCGCTCAATGCAACCGCGACATTTTTCAGTCTTAGTAA 3792369
QY	337	TrpThrArgIleTrpAspTrpThrSer	345
DB	3792370	TTAGTGGCCCGCGCTTTGGTTCACCGCGGCGCTCGCGGGAACACACAGACGTGTATGCCG	3792429
QY	346	TyrProAsn	ArgSerLeuArgTyrValLeuAspIleSerAl 359
DB	3792430	TGATCGCGCATACCTTCGACCCATTGAAGGAGAGACGACCATGTCTTTGTGA-TGCG	3792488
QY	359	agluPro	TrpLeuThrPheGlyValGln 368

DB	3792489	AACCCGAGATGCTGGCAGCGCGCGCATCCGATTGGCCCGGATCCGCTCGGCGATCAGC	3792548
QY	369	ProAsnProProValProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAs	388
DB	3792549	GCCGCGACCGCGCGCGCGCGCGCGCATCCAG-GTTGCCGCGCGCGCGCGCGCA	3792607
QY	388	pProPheAsnSerAspArgMetLeuTyrGly	398
DB	3792608	GGTGTCTGCTGCGCTCTCGCGCTGTTTGGCAGCACGCCCGGCTATCAGCGGCTCAG	3792667
QY	399	ThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnI	418
DB	3792668	CGCCAGCGCAGCATCTTCAGCACCGCTGTCGAGCGCTGCTCCGCGCGCAACT	3792727
QY	418	eHisIleAlaProMetValLysGlyLeuGluGluThrAlaValAsnAspLeuIleSerPr	438
DB	3792728	GTATGCGCGCGCGAGAGCACACCGTGCAGCATGTGTCTCAACGGCATCAACGCGC	3792787
QY	438	oProSer	GlyAlaProLeuIleSe 446
DB	3792788	CACCGACACTGTTCGCGCGCGCTGATCGCGCAGCGCGCAACCGGAGCGGAGAA	3792847
QY	446	rAlaLeuGlyAspLeuGlyPhe-ThrHisAlaAsp	ValThrAlaValProSerTr 465
DB	3792848	CCCGACCGGCAAAACCGCGCGCTGCTGTTCGCAACCGCGCGCAACCGCTTACCCAGAC	3792907
QY	465	hrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeuAsnPr	485
DB	3792908	GACCGCGCGGTGGCGCGCAACCGCGCGCGCGGTGATCGCGCAACCGCGCGGGC	3792967
QY	485	roSerIleIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArgHisV	505
DB	3792968	CGCGCGCGCGCGCGG-GCGCGCGC	3792992
QY	505	alAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySer	GluProGlyGlyV 524
DB	3792993	CGCGCGCGCTCGCGCGCAACCGCGGTGCTGTACGGCAACCGCGCGCGCGCGCA	3793050
QY	524	alThrThrGlyGlyThr	ValAlaAlaSerAlaAspGlySerA 538
DB	3793051	TCGGCGCGCGCGCACCGGAACCGTGTGTACCGCGCGCGCGCGCGCGCGCGCGG	3793110
QY	538	rgPheValTrp-AlaProGlyAspProGlyGlnProValValTyrAlaValGly	PheG 557
DB	3793111	CCTGGCTGTGGGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGTGTTCG	3793170
QY	557	lyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgV	577
DB	3793171	GCACGCGCGCGCGCGCGCACCGCGCGCAACCGCGCGCGCTTTAACACG	3793223
QY	577	alAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyV	597
DB	3793224	TTGACTTCTTCGTCGCGCG	3793242
QY	597	lyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPr	617
DB	3793243	CG	3793301
QY	617	heHisAlaValProGlyLysGlyGlyLeuTrpLeuAlaAlaSerSerGlyLeuTyrH	637
DB	3793302	GGCGCATCGCGCGCAAAACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	3793359
QY	637	isSerThrAsnGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnV	657
DB	3793360	CGCGCGCGCGCGCGCGCGCGGTGTCTGTGCGCGATGGCGCGCGCGCGCGCGCGCG	3793419
QY	657	alGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValGlyThrI	677
DB	3793420	GGCGCGCGCGGAAGAAATTCAGCGGTGGCGC	ACCC 3793455
QY	677	leGlyGlyValThrGlyAlaTyrArgSerAspCysGlyThrTrpValLeuIleA	697
DB	3793456	TCACCGCGCGCGCACCGGAGGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG	3793508



QY 505 alAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySer---GluProGlyGlyV 524  
Db 3802171 --GCCGGCGCGCTCGCGCGCAACCGCGGTGTGTACGGCAACCGCGCGCGCGCA 3802228  
QY 524 alThrThrGlyGlyThr-----ValAlaAlaSerAlaAspGlySerA 538  
Db 3802229 TCGGGGGCGCGGCGCACCGGAACCGGTGTACCGCGGGCGCGCGCGCGCGCGG 3802288  
QY 538 rgPheValTrp-AlaProGlyAspProGlyGlnProValValTrpAlaValGly--Phe 557  
Db 3802289 CCTGGCTGTGGGCGACCGCGCGGGCGCGCGGAGCGCGGTGACGGCGGTGTGTTCG 3802348  
QY 557 lyAsnSerTrpAlaAlaSerGlnGlyValProAlaAlaGlnIleArgSerAspArgV 577  
Db 3802349 GCGACGGCGGGCGCGCGCACCGCGCGCAACCGCGCGCGCGGTTTAAACAGC----- 3802401  
QY 577 alAsnProIysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyG 597  
Db 3802402 -----TTGACCTCTTTCGGTTCGGCG 3802420  
QY 597 lyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetP 617  
Db 3802421 GCGCGCGCGCGCGGTGGCGACCGCGCGGTGTTCGGCGCGCGCGCGGACCGCGCGGACC- 3802479  
QY 617 heHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrH 637  
Db 3802480 --GGCGGATCGCGCGGCAAAACACGAGACCGCGCGCGCGCGCGCGCGCGCGCG 3802537  
QY 637 isSerThrAsnGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnV 657  
Db 3802538 GCGCGCGCGGTGGCGCGCGGTGTTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3802597  
QY 657 alGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValGlyThrI 677  
Db 3802598 GGGCGCGCGGAGATTCCAGCGGTGGCGCC-----ACCC 3802633  
QY 677 leGlyGlyValThrGlyAlaTyrArgSerAspCysGlyThrThrTrpValLeuIleA 697  
Db 3802634 TCACCGGGGCGCACCGGAGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC- 3802686  
QY 697 snAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnL 717  
Db 3802687 -----TACGGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3802729  
QY 717 euArgArgValTyrIleGlyThrAsnGlyArgGly----- 728  
Db 3802730 TCACACCGCGCGGTGTGTCCACCGCGCGCGCGCGCGCGGTACCGCGCGCGCGCGCTCTG 3802789  
QY 729 -----IleValTyrGlyAspIleGlyAlaProSerGly 740  
Db 3802790 GAGCGTGGTGTACGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3802830

## RESULT 11

US-09-252-991A-13875/c  
; Sequence 13875, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 13875  
; LENGTH: 2319  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-13875

Alignment Scores:  
Pred. No.: 3.35e-05 Length: 2319  
Score: 164.00 Matches: 169  
Percent Similarity: 36.56% Conservative: 80  
Best Local Similarity: 24.82% Mismatches: 275  
Query Match: 4.06% Indels: 160  
DB: 40 Gaps: 40  
US-09-917-376-3 (1-740) x US-09-252-991A-13875 (1-2319)  
QY 72 AlaAlaAspProIleAsnThrAsn-----LysValTrpAlaAlaValGlyMetTyrThr 89  
Db 1837 GCAGCGGTACCGCAAGCGCGCAACCGCGCTGACGGTGGCGTTCATACCGACCGCGCAG 1778  
QY 90 AsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrp 109  
Db 1777 GCACGCGGACACCA-----CCGTGG 1757  
QY 110 GlnIleThrProLeuProPheLysLeuGlyGlyAsnMet-----ProGly 124  
Db 1756 TG-GTCGGCCCC-----GGCGGACGTTTCGAGGTTCCGCTGAACCGCGCG 1713  
QY 125 ArgGlyMetGlyGluArgLeuAlaVal-----AspPro-----AsnAsnAspAsn 139  
Db 1712 CTGACCAATGCGGACGCGTACGCTGATCGTTACCGACCGCGCGCGCAACACAGCACC 1653  
QY 140 IleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAsp-----Ser 157  
Db 1652 CCGGTACCGTCCAGGCGCGCGACACACAGCCCCCGCGCGCGCGCGCGCGCGCGGTG 1593  
QY 158 GlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnPro 177  
Db 1592 GCGCGGACGCGCAGCGTCCCGCAACGACAGCGCGCGCGCGCGCGCGCGCGCGTGTGAC 1533  
QY 178 ThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAla-----Phe 195  
Db 1532 ACCGATGCGCAGCGC---CAGCGCGACACACCGGTGTGTGTGCGCGCGCGCGCGCGTTC 1476  
QY 196 AspLysSerSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAsp 215  
Db 1475 GAGTCCCGTCCGACCGCGCGCTGACCAATGGCGAGCGGTGACGCGTATCGTTTACCGAC 1416  
QY 216 Pro-----AsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGln 231  
Db 1415 CCGCGCGGCAACGACGACCGCGCTC-----ACCGCGCA 1380  
QY 232 Ala-----ValProGlyAlaProThrGlyPheIleProHisLysGlyValPheAspPro 249  
Db 1379 GCTCCCGACTTCCCGACGCGCGCGCGTCAATGCCAGCAACGCGCGGTCTCTCAGCGGT 1320  
QY 250 ValAsnHis-----ValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAspGly 267  
Db 1319 ACGCGGGAAGCGCGGTGACCATCGATACCGGACGCGCAACCGATCCGATCCGCGCGAG 1260  
QY 268 SerSerGlyAspVal-----TrpLysPheSer-----ValThrSerGly 280  
Db 1259 ACCAGCGCGCATGCCAACGCGCAACTGGAGCTTACCCCGCGGTAGCAACTGCGCGGTGCG 1200  
QY 281 ThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyr 300  
Db 1199 ACC-----GTGGTCAATGTGTGGCGGACGCGCGCGCGCAACAGCAGCGCGCGG 1149  
QY 301 SerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetVal-----Ala 316  
Db 1148 ACCTCCATACCGTTCGACGCGGTGGCGCGCGCGCGCGGTGGTGGCGCGCGAGCAACGCGC 1089  
QY 317 ThrGlnIleSerTrpTrp-----ProAspThrIleIlePheArgSerThrAspGlyGly 334  
Db 1088 AGCGAATCAGCGGACTGCGCAACCGCGCGCGCGCGCGGTG---ACCTGACCGCGCGCAAT 1032  
QY 335 -----AlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsn 348



1031	Db	GGCAATCGATCGGCCAGACACACCGCCGATGCCAACGGCAACTGGTCTTTTCACGCCCGTC	977
349	Qy	ArgSerLeu-ArgTyrValIleuAspIleSerAlaGluProTrpLeuThrPhe-----G1	366
971	Db	ACCCCGCTGCCGACGGTA-----CCGTGGTCAACGTGGTGCCAGG	930
366	Qy	yValGlnProAsnProProValProSerProLysLeuGlyTrpMetAspGluAlaMetAl	386
929	Db	GATCCGCCCGGCAACACAGCAGTCCCGCCGCGACAGCGTTACCGTGGATCGCGTC-GCGCCGCG	871
386	Qy	a-----IleAspProPheAsnSerAspArgMetLeuTyrGlyThr-----G1	400
870	Db	CACGCCCACCGCTCATCCGACCAACGGT---ACGACCCTCAGCGGCACCGCGAGCCGGG	814
400	Qy	yAlaThrLeuTyrAlaThrAsn-----AspLeuThrLysTrpAs	413
813	Db	CGCTACCGTGACCTGACCGAGGCAACGGCAACCCGATTGGCCAGGTCAACGGCC---GA	757
413	Qy	pSerGlyGlyClnIleHisIleAlaProMetValLysGlyLeuGluThrAlaValas	433
756	Db	CGGCAGCGGCAACTGGACCTTCACCCCGAGCAGCCGCTTGCCCAACGGCACCGTGTCAA	697
433	Qy	nAspLeuIleSerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyG1	453
696	Db	CGCCACCGGCTACCGACCCGCTCCGGC-----AACGC	667
453	Qy	yPheThrHisAlaAspValThr-----AlaValProSerThrIlePheThrSerPr	470
666	Db	CAGTTCGCGCGCAGCGTCAACGTGGACCCCGTGGCACCGGCCACCGCAGTGTCACACC	607
470	Qy	oValPheThrThrGlyThrSerValAsp---TyrAlaGluLeuAsnProSerIleVal	489
606	Db	G-----AGCAACGGCACCAACGCTCAGCGGCACCGCGAGCGCGCGCACCGTCACCT	553
489	Qy	lArgAlaGlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerTh	509
552	Db	GGCCGATGGCAACGGCAATCCCATCGGGCAG-----GTCAACGC	514
509	Qy	rAspGlyGlyLysAsnTrp---PheGlnGlySerGluProGlyGlyValThrThrGlyG1	528
513	Db	CGATGGCAGGGCAACTGGAGCTTCACTCCGACCAACGGCG-----TTGCCCAACGG	463
528	Qy	yThrVal-----AlaAlaSerAlaAspGlySerArgPheValTrpAlaProGlyAspPr	546
462	Db	CACCGTGTGTCAACGCCACCGGCCACCGACGCTCC-----429	
546	Qy	oGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyVa	566
428	Db	-----GGCAACACCATGTGGCGGCAGCAGTGTCTAC	400
566	Qy	lProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrAlaLeuSe	586
399	Db	CGTGGCATCGGTAGCCCGGCACCGCATGATCAACCCACAGCAACGGCACCCAGCTCAG	340
586	Qy	rAsn-----GlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnPr	602
339	Db	CGGCACCCCGAGCGCGGCACGCGTACCTGACCGATGGC-----AACGGCAACCC	286
602	Qy	oVal-AlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaValProG	622
285	Db	GATCGGCAGGTCAACCGCGACCGCAGCGGCAACTGGAGCTTCACC-CCGTCCACCGCGC	227
622	Qy	lLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsnGlyG	642
226	Db	TGGCGGATGGAAACCGTGTCAACGCCACCGGCACCGATCCCGGGGGCAACACCGCGCC	167
642	Qy	lySerSerTrpSerAlaIleThrGlyVal-----SerSerAlaValAsnValGlyPr	659
166	Db	AGGCAGC---ACCACCGTCGATGGCTGGCGCCGACCCACCGCCGCTCAACCTGACGA	110
659	Qy	heGlyLysSerAlaProGlySerSerTyrProAlaValPheValValGlyThrIleGly	678
109	Db	ACGGCAGCAGCTCAGCGGCACCTCGGAAACCGGGCAGCACCGGTGATCTCTCACGACGG	51

RESULT 12  
 US-07-706-699-4  
 ; Sequence 4, Application US/07706699  
 ; Patent No. 5204254  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Schmid; Candusio; B ck  
 ; TITLE OF INVENTION: Maltopentaose  
 ; TITLE OF INVENTION: Producing Amylaase  
 ; NUMBER OF SEQUENCES: 4  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Collard, Roe & Galgano, P.C.  
 ; STREET: 1077 No. 5204254thern Boulevard  
 ; CITY: Roslyn  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 11576  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Wordperfect 5.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07706.699  
 ; FILING DATE: 19910529  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: GR 40 17 595.2  
 ; FILING DATE: 31 MAY 1990  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Allison C. Collard  
 ; REGISTRATION NUMBER: 22,532  
 ; REFERENCE/DOCKET NUMBER: SCHMID ET AL-W  
 ; NAME: Thomas M. Galgano  
 ; REGISTRATION NUMBER: 27,638  
 ; REFERENCE/DOCKET NUMBER: SCHMID ET AL-W  
 ; NAME: Edward R. Freedman  
 ; REGISTRATION NUMBER: 26,048  
 ; REFERENCE/DOCKET INFORMATION:  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 516-365-9802  
 ; INFORMATION FOR SEQ ID NO: 4:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 5741 bases  
 ; TYPE: NUCLEIC ACID  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; US-07-706-699-4

Alignment Scores:		
Pred. NO.:	0.000168	5741
Score:	163.00	Matches: 177
Percent Similarity:	29.10%	Conservative: 98
Best Local Similarity:	18.73%	Mismatches: 276
Query Match:	4.04%	Indels: 394
DB:	1	Gaps: 47

US-09-917-376-3 (1-740) x US-07-706-699-4 (1-5741)

Qy	47	AlaAsnGlyArgTrpIlePro-----LeuLeuAspTrpValGly	59
Db	2562	TCITTCAGGAACAAATTCACAGGTCCAAATGAACACAGGATCTTCATTGATCAGATTGGT	2621
Qy	60	TrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIle-----	76
Db	2622	TGG-----TACGATGGCGTAAAGTGGCTTGATTCAGATCCTTTTGAACGGGAA	2669
Qy	77	-----AsnThrAsnLysValTrpAlaAlaValGlyMet	87
Db	2670	CCTTAAGAGCGCTCGGCAACACCTTAAGACCTTAAGTGTGTTAATGTAACCTGAACACTACT	2729
Qy	88	TyrThrAsnSerTrpAspProAsnAspGly-----AlaIleLeuArgSer	102



```
QY 577 ValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGly 596
Db 4743 -----GGCTCATCAGCTGGTACCACAATTAGCCAGATGATATG 4781
QY 597 GlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMet 616
Db 4782 CCACAAACCAATT-----GCAATTGATACCATCCAGAGCAAGGTGCAACA----- 4826
QY 617 PheHisAlaValProGlyLysGlyGly-----AspLeuTrp----- 628
Db 4827 ----CATGATATGTGGGGGAAAAAATGGTGTGATCATCTTGGGAGGACAGATTACCT 4883
QY 629 -----LeuAlaAlaSerSerGlyLeuTyrHisSer----- 638
Db 4884 GATTACCAACATAATATCGCATCTAATATGTTCCATTGAGGCTATATTTCTAGACAGTT 4943
QY 639 -----ThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSer 652
Db 4944 GATGGTGTATTTCCTGTTGACGATGGAGGAATAAATTATATA-----ACGGGTGAGGAA 4997
QY 653 SerAlaValAlaValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPhe 672
Db 4998 GCAGGAATTACAGTAAAGTTTCTAAA-----GGTAAAGGTAT----- 5036
QY 673 ValValGlyThrIleGlyGlyValThrGlyAlaTyrArgSerAspCysGlyThrThr 692
Db 5037 -----TCAACATTTGTGGGGGTGTAGATGCT-----GATGATGCAGTTGATCCT 5081
QY 693 TrpValLeuIleAsn 697
Db 5082 AGTAAACTTGTGAAC 5096
```

## RESULT 13

```
US-07-998-931-4
; Sequence 4, Application US/07998931
; Patent No. 5304723
; GENERAL INFORMATION:
; APPLICANT: Schmid; Candussio; Bck
; TITLE OF INVENTION: Malopentaoose Producing Amylases
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESSES:
; ADDRESS: Collard, Roe & Galgano, P.C.
; STREET: 1077 No. 5304723thern Boulevard
; CITY: Roslyn
; STATE: New York
; COUNTRY: USA
; ZIP: 11576
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 5.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/998,931
; FILING DATE: 19921229
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GR 40 17 595.2
; FILING DATE: 31 MAY 1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Allison C. Collard
; REGISTRATION NUMBER: 22,532
; REFERENCE/DOCKET NUMBER: SCHMID ET AL-W2
; NAME: Thomas M. Galgano
; REGISTRATION NUMBER: 27,638
; REFERENCE/DOCKET NUMBER: SCHMID ET AL-W2
; NAME: Edward R. Freedman
; REGISTRATION NUMBER: 26,048
; REFERENCE/DOCKET NUMBER: SCHMID ET AL-W2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 516-365-9802
; INFORMATION FOR SEQ ID NO: 4:
```

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5741 bases
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-07-998-931-4

Alignment Scores:
Pred. No.: 0.000168 Length: 5741
Score: 163.00 Matches: 177
Percent Similarity: 29.10% Conservative: 98
Best Local Similarity: 18.73% Mismatches: 276
Query Match: 4.04% Indels: 394
DB: 1 Gaps: 47

US-09-917-376-3 (1-740) x US-07-998-931-4 (1-5741)
QY 47 AlaAsnGlyArgTrpIlePro-----LeuLeuAspTrpValGly 59
Db 2562 TCTTCAGGAAACAAATTCAGGTCCAAATGAACAGGATTTCTTCATTGATCAGATTGGT 2621
QY 60 TrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIle----- 76
Db 2622 TGG-----TACGATGCGGTAAAGTGGCTGTGATTGATTCAGATCCTTTTGAAGGAA 2669
QY 77 -----AsnThrAsnLysValTrpAlaAlaValGlyMet 87
Db 2670 CCTAAGAGCCTGCGCAACACCTAAGACCTAAGTGTGTTAATGTTAACTGAAACTACT 2729
QY 88 TyrThrAsnSerTrpAspProAsnAspGly-----AlaIleLeuArgSer 102
Db 2730 GTAACATTTGAGTGGGCAACATCTGATGTTATGTCGTTGAATACGAGATTTTACGTGAT 2789
QY 103 SerAspGlnGlyAla----- 107
Db 2790 GAGGATGTTGTTGCTTCAACTATTCGTACAACTATTCGCGATGAAGACCTTTAATCCAGAT 2849
QY 108 ---ThrTrpGlnIleThrProLeuProPheLysLeuGlyGly-----AsnMetProGly 124
Db 2850 ACAACCTACACTTATTTCTGCTGAGCTGTTGGAGAGCGCGGAGAAATCCGCCCCAGT 2909
QY 125 ArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGly 144
Db 2910 GAAGCGTTA---AAAGTGACTACATTAGAAGAAATGATGAACCTACAGGAACCGCTGAG 2966
QY 145 AlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThr-----TrpSer 162
Db 2967 GCGCCAGAAAAATTACGTATAGCTGATATACAGATACACAGTTACAATCAACTGGAAT 3026
QY 163 GlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGly 182
Db 3027 GCATCTTAATGGTTAC-----GTAAACAGGA 3050
QY 183 TyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerLeu 202
Db 3051 TATGAGGTTCTGCTGATGTTGGTT-----ATT 3080
QY 203 GlyGlnAlaSerLysThrIlePheValGlyVal----- 213
Db 3081 GCGGAAACAAACAGGACAACATTTCATAGATACTGATTAGATGCTGATAGGACCTATACG 3140
QY 214 -----AlaAspPro----- 216
Db 3141 TATACGATTGTTGCTCTCGGAGATGGGGCAAAAGTCTGATCCGAGCAAGCGTTAGAA 3200
QY 216 ----- 216
Db 3201 GTAACAACTCAAGAAAAACAGAGGAATCTAGTAACTATATAATAAAAGGCTTT 3260
QY 217 AsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGly--- 235
Db 3261 GATACCCCATATATGCAATTATCGTCCGGAAGGTGGAGAGTGGACGATGCTGTTCCAGGAATT 3320
```

QY 235 ----- 235  
Db 3321 AGAATGGAAGAAATCAGAAATAGCAGGCTATAGTAAGTTAAACCGTTGATATTCGGGAAGCA 3380  
QY 236 -----AlaProThrGlyPheIleProHisLysGlyValPhe-----AspProValAsn 251  
Db 3381 AGCAAGTTGGAAGTAGACCTTTAATAATGACGCTGGGCTTGGGATAGTGATCAAGAGAAT 3440  
QY 252 HisValLeuTyr----- 255  
Db 3441 AATTATTATTGAGCCAGCGTGTTCATACGTACATTTCCAGTCATCAAGGAAGAGAGAGAG 3500  
QY 256 IleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrp----- 273  
Db 3501 ATTATTCCAGGTAACACGAGGAGCACCACATCGATGGTAAATAAGTGACGATTTACTATCAA 3560  
QY 274 -----LysPheSerValThrSerGlyThrTrpThrArgIle 285  
Db 3561 AATGCTTTGATACGCGGTATGTTTCATTACCGCCCAAGAGCGGAAATTTGGACCAACGCC 3620  
QY 286 SerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIle 305  
Db 3621 CCAGATTAAATGGAAGAT-----TCAGATTTCGAAGTTATAGTAGGTTAACGCTT 3674  
QY 306 AspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrpProAspThr 325  
Db 3675 GATATTGGTGAAGCTAATCGTCAGAGTGGCT----- 3707  
QY 326 IleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIleTrpAspTrpThrSer 345  
Db 3708 -----TTCAATAACGAGCGCGC-----CTTTGGATAGTGATAAT 3743  
QY 346 TyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpLeu----- 363  
Db 3744 GAAATAAAT-----TATTTCTCAATATTGGCGTAACACTTATATATACAGGA 3791  
QY 364 -----ThrPheGlyValGlnProAsnProProValProSer----- 375  
Db 3792 AAAAAGCGTTACGTGGAGAGATTATGAGGTAAGCCAAAGACCACCAATTAGTAGGAAT 3851  
QY 376 -----ProLys 377  
Db 3852 GAAGTAATCATTTATTATAAAATGTTTGTGATACACCGTATGTTCAATTATCGTCCAGAA 3911  
QY 378 LeuGly-----Trp-----MetAspGluAla----- 384  
Db 3912 GGTGGTACGTGGACAAATGCACGAGAAATAAATGGATAAGTCAGAAATAGCAGGTTAC 3971  
QY 385 -----MetAlaIleAspProPheAsnSerAspArgMetLeuTyr-----Gly 398  
Db 3972 AGTAAATAACGCTTGATATGTCGCGCAGATCGAGTAGAGTAGCCCTTTAATGACGCT 4031  
QY 399 ThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGly----- 415  
Db 4032 CGTGGTGCA-----TGGGATAGTATAACGAACGCT 4061  
QY 416 -----GlyGlnIleHisIleAlaProMetValLysGlyLeuGlu 429  
Db 4062 AATTATCTTTGTAGTCGGTAACAACTATTATGACCAAGGAATTAACCGCGCACCTGCT 4121  
QY 430 ThrAlaValAsnAspLeuSerProProSerGlyAlaProLeuIleSerAlaLeuGly 449  
Db 4122 CAGGTGAACATCGCGTGTTCCTGATGATGAGAGATCCG-----GGA 4166  
QY 450 AspLeuGlyGlyPheThrHis-----AlaAspValThrAlaValPro 463  
Db 4167 GATATTGAAGACCTCATCATACCTCCCTTCAAAGCCGACTGATTAAACAGCAATAGCT 4226  
QY 464 SerThrIlePheThrSerProValPheThrGlyThrSerValAspTyrAlaGluLeu 483  
Db 4227 ACTGCTCATACTGTTTTCATTAAAGCTGGACA-----GCTTCAGCAGCAGCATGAGAATGA 4280

QY 484 AsnProSerIleIleValArgAlaGly----- 492  
Db 4281 GCTGGGTACAAAATTTATCGAGATGGTGTGAAATCGGTGTACTGAATCAACAACCTTAT 4340  
QY 493 -----SerPheAspProSerSerGlnProAsnAspArg 503  
Db 4341 ACGGATTCAGGTTAACGGCAGAAACACGTATATAGCTATATGGTACAACTTATGATACT 4400  
QY 504 HisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGly 523  
Db 4401 TCTAATAATTTCTCGGCATTAAGTGAATGACAAATGAAACCCGCGGAAAAACGGT 4460  
QY 524 ValThrThrGlyGlyThrValAlaIleAspGlySer-----ArgPheValTrp 541  
Db 4461 GTTGATCCAGGAGGGATATGCTTATTCACGAATCCATCGTTGTTGTAAGAAGTAACA 4520  
QY 542 AlaPro-----GlyAspProGlnProValValTrpAla 553  
Db 4521 AGCCCAATCACAAATGATGGTGTAAATGACGGGAATGGACAGATGATATGTTGATGCA 4580  
QY 554 ValGly-----PheGlyAsnSerTrpAla----- 561  
Db 4581 ATTGGTATGGTGGTACGACCCACCGTTCGCTCGGGGACAAATGCTCTATCATCAAGAA 4640  
QY 562 -----AlaSer 563  
Db 4641 CCAATGGACCTTACTCACCTATGGGAGCATGGACCATGAGTACTTGTATCTTGTGG 4700  
QY 564 GlnGlyVal-----ProAlaAsnAlaGlnIleArgSerAspArg 576  
Db 4701 CAATATGTAGATGAACAGATATATTATGACCCAGCTAACGCA----- 4742  
QY 577 ValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGly 596  
Db 4743 -----GGCTCATCAGCTGTTACCAATAGCCACCATGATGATGATG 4781  
QY 597 GlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMet 616  
Db 4782 CCACAAACCAATT-----GCAATTGTATACCATCCAGAGCAAGGTGCAACA----- 4826  
QY 617 PheHisAlaValProGlyLysGluGly-----AspLeuTrp----- 628  
Db 4827 ---CATGATATGCGGGGAAAAATGGTGGTGAATCACTTTGGGAGGAGCCAGATTTACCT 4883  
QY 629 -----LeuAlaAlaSerSerGlyLeuTyrHisSer----- 638  
Db 4884 GATTACCACTAAATATCGCATCTAATATGTTCCATTCAGCTATATTTCTAGACAGCTT 4943  
QY 639 -----ThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSer 652  
Db 4944 GATGGTGTATTTCTCTGTGACGATGGAGGAATAAATTATAAA-----ACGGGTGAGGAA 4997  
QY 653 SerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPhe 672  
Db 4998 GCAGGAATACAGTAAGTTTCTTAA-----GGTAAAGGGTAT----- 5036  
QY 673 ValValGlyThrIleGlyGlyValThrGlyValTyrArgSerAspCysGlyThrThr 692  
Db 5037 -----TCACATATGCGGGGGTGTAGATGCT-----GATGATGCAAGTTGATCCT 5081  
QY 693 TrpValLeuIleAsn 697  
Db 5082 AGTAAACTTGTGAAC 5096

## RESULT 14

US-09-679-279-1  
; Sequence 1, Application US/09679279  
; Patent No. 6524841  
; GENERAL INFORMATION:  
; APPLICANT: McDaniel, Robert  
; APPLICANT: Volchegursky, Yanina  
; TITLE OF INVENTION: Recombinant Megalomicin Biosynthetic  
; TITLE OF INVENTION: Genes and Uses Thereof

```
FILE REFERENCE: 300622004700
CURRENT APPLICATION NUMBER: US/09/679,279
CURRENT FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/158,305
PRIOR FILING DATE: 1999-10-08
PRIOR APPLICATION NUMBER: US 60/190,024
PRIOR FILING DATE: 2000-03-17
NUMBER OF SEQ ID NOS: 34
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 47981
TYPE: DNA
ORGANISM: Micromonospora megalomicea
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(144)
OTHER INFORMATION: megBVI (megT), TDP-4-keto-6-deoxyglucose-2,3-dehydratase;
OTHER INFORMATION: SEQ ID NO: 2= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (928)...(2061)
OTHER INFORMATION: megDVI, TDP-4-keto-6-deoxyglucose 3,4-isomerase;
OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 3,4-isomerase;
OTHER INFORMATION: SEQ ID NO: 3= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (2072)...(3382)
OTHER INFORMATION: megDI, rhodosaminyl transferase (eryCIII homolog);
OTHER INFORMATION: TDP-megosamine glycosyltransferase;
OTHER INFORMATION: SEQ ID NO: 4= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (3462)...(4634)
OTHER INFORMATION: megG (megY), mycarosyl acyltransferase, mycarose O-acyltransferase
OTHER INFORMATION: SEQ ID NO: 5= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (4651)...(5775)
OTHER INFORMATION: megDI, deoxysugar transaminase (eryCI, DnrJ homolog);
OTHER INFORMATION: TDP-3-keto-6-deoxyhexose 3-aminotransaminase;
OTHER INFORMATION: SEQ ID NO: 6= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (5822)...(6595)
OTHER INFORMATION: megDII, daunosaminyl-N,N-dimethyltransferase (eryCVI homolog);
OTHER INFORMATION: SEQ ID NO: 7= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (6592)...(7197)
OTHER INFORMATION: megDV, TDP-4-keto-6-deoxyglucose 3,5-epimerase (eryBVII, dnmU homolog);
OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 3,5-epimerase;
OTHER INFORMATION: SEQ ID NO: 8= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (7220)...(8206)
OTHER INFORMATION: megDV, TDP-hexose 4-ketoreductase (eryBIV, dnmV homolog);
OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 4-ketoreductase;
OTHER INFORMATION: SEQ ID NO: 9= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (8228)...(9220)
OTHER INFORMATION: megBII-1 (megDVII), TDP-4-keto-L-6-deoxy-hexose 2,3-reductase;
OTHER INFORMATION: SEQ ID NO: 10= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (9226)...(10479)
OTHER INFORMATION: megBV, mycarosyl transferase, mycarose glycosyltransferase;
OTHER INFORMATION: SEQ ID NO: 11= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (10483)...(11424)
OTHER INFORMATION: megBIV, TDP-hexose 4-ketoreductase;
OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 4-ketoreductase;
OTHER INFORMATION: SEQ ID NO: 12= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (12181)...(22821)
OTHER INFORMATION: megAI; SEQ ID NO: 13= translated amino acid sequence
NAME/KEY: misc feature
LOCATION: (12505)...(13470)
OTHER INFORMATION: megAI, AT-L
NAME/KEY: misc feature
LOCATION: (13576)...(13791)
OTHER INFORMATION: megAI, ACP-L
NAME/KEY: misc feature
LOCATION: (13849)...(15126)
OTHER INFORMATION: megAI, KS1
NAME/KEY: misc feature
LOCATION: (15427)...(16476)
OTHER INFORMATION: megAI, AT1
NAME/KEY: misc feature
LOCATION: (17155)...(17694)
OTHER INFORMATION: megAI, KR1
NAME/KEY: misc feature
LOCATION: (17947)...(18207)
OTHER INFORMATION: megAI, ACP1
NAME/KEY: misc feature
LOCATION: (18268)...(19548)
OTHER INFORMATION: megAI, KS2
NAME/KEY: misc feature
LOCATION: (19876)...(20910)
OTHER INFORMATION: megAI, AT2
NAME/KEY: misc feature
LOCATION: (21517)...(22053)
OTHER INFORMATION: megAI, KR2
NAME/KEY: misc feature
LOCATION: (22318)...(22575)
OTHER INFORMATION: megAI, ACP2
NAME/KEY: CDS
LOCATION: (22867)...(33555)
OTHER INFORMATION: megAI; SEQ ID NO: 14= translated amino acid sequence
NAME/KEY: misc feature
LOCATION: (22957)...(24237)
OTHER INFORMATION: megAI, KS3
NAME/KEY: misc feature
LOCATION: (24544)...(25581)
OTHER INFORMATION: megAI, AT3
NAME/KEY: misc feature
LOCATION: (26998)...(27258)
OTHER INFORMATION: megAI, ACP3
NAME/KEY: misc feature
LOCATION: (27393)...(28590)
OTHER INFORMATION: megAI, KS4
NAME/KEY: misc feature
LOCATION: (28897)...(29931)
OTHER INFORMATION: megAI, AT4
NAME/KEY: misc feature
LOCATION: (29953)...(30477)
OTHER INFORMATION: megAI, DH4
NAME/KEY: misc feature
LOCATION: (31396)...(32244)
OTHER INFORMATION: megAI, ER4
NAME/KEY: misc feature
LOCATION: (32257)...(32799)
OTHER INFORMATION: megAI, KR4
NAME/KEY: misc feature
LOCATION: (33052)...(33312)
OTHER INFORMATION: megAI, ACP4
NAME/KEY: CDS
LOCATION: (33666)...(43271)
OTHER INFORMATION: megAI; SEQ ID NO: 15= translated amino acid sequence
NAME/KEY: misc feature
LOCATION: (33780)...(35027)
OTHER INFORMATION: megAI, KS5
NAME/KEY: misc feature
LOCATION: (35385)...(36419)
OTHER INFORMATION: megAI, AT5
NAME/KEY: misc feature
LOCATION: (37069)...(37604)
OTHER INFORMATION: megAI, KR5
NAME/KEY: misc feature
LOCATION: (37860)...(38120)
OTHER INFORMATION: megAI, ACP5
NAME/KEY: misc feature
```

```

; LOCATION: (38187)...(39470)
; OTHER INFORMATION: megAIII, KS6
; NAME/KEY: misc feature
; LOCATION: (39795)...(40811)
; OTHER INFORMATION: megAIII, AT6
; NAME/KEY: misc feature
; LOCATION: (41406)...(41936)
; OTHER INFORMATION: megAIII, KR6
; NAME/KEY: misc feature
; LOCATION: (42168)...(42425)
; OTHER INFORMATION: megAIII, ACP6
; NAME/KEY: misc feature
; LOCATION: (42585)...(43271)
; OTHER INFORMATION: megAIII, TE
; NAME/KEY: CDS
; LOCATION: (43268)...(44344)
; OTHER INFORMATION: megCII, TDP-4-keto-6-deoxyglucose 3,4-isomerase;
; OTHER INFORMATION: SEQ ID NO: 16= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (44355)...(45623)
; OTHER INFORMATION: megCIII, desosaminyl transferase, desosamine glycosyltransferase;
; OTHER INFORMATION: SEQ ID NO: 17= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (45620)...(46591)
; OTHER INFORMATION: megBII-2(megBII), TDP-4-keto-6-deoxy-L-glucose 2,3 dehydratase,
; OTHER INFORMATION: TDP-4-keto-6-deoxyglucose 2,3 dehydratase;
; OTHER INFORMATION: SEQ ID NO: 18= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (46660)...(47403)
; OTHER INFORMATION: megH, TEII; SEQ ID NO: 19= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (47411)...(47980)
; OTHER INFORMATION: megF, C-6 hydroxylase; SEQ ID NO: 20= translated amino acid sequence
; US-09-679-279-1

```

## Alignment Scores:

```

Pred. No.: 0.00757 Length: 47981
Score: 160.50 Matches: 205
Percent Similarity: 28.61% Conservative: 78
Best Local Similarity: 20.73% Mismatches: 334
Query Match: 3.98% Indels: 374
DB: 4 Gaps: 45

```

US-09-917-376-3 (1-740) x US-09-679-279-1 (1-47981)

```

QY 15 GlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrVal 34
Db 16683 GGTGGTGAGCCCGGACGGG-----CCGGGTGCACCTGTGCGCGGCTCGA 16730
QY 35 ArgThrAspIleGly-----MetTyrArgTyrAspAlaAlaAsnGlyArg 50
Db 16731 CGCGGTGGTCCGGTGGCGGCGTCTCTGCTGTTGCTGTCGCGGAGCCGCGGCCGA 16790
QY 51 TrpIle-ProLeuLeuAspTyrValGlyTyrAsnAsnTyrGlyTyrAsnGlyValValse 70
Db 16791 ACATCCCGGCTG-----CGCGGTGAC 16811
QY 70 rIleAlaAlaAspProIleAsnThrAsnLysValTyrAlaAlaValGlyMetTyrThrAs 90
Db 16812 GTCTGTGTCGACACCTCGACCTGACCTGACCCAGCGGTGGCGCGGTGCGGAGTGTTC 16871
QY 90 nSerTyrAspProAsnAspGlyAlaIle-----LeuArgSerSe 103
Db 16872 GATCTGGGTGGTCACCGAGACCGCTGCGCGCTCGGCGCCCTTCGAACGGCTCCGCGACCC 16931
QY 103 rAspGingIlyAlaThrTrp----- 109
Db 16932 GGCCACCGGCGCTCTGGGCCCTCGTGGGTGCTGCGGTGCTGCCCTGAGAACCCCGCGTCTG 16991
QY 110 -----GlnIleThrProLeuProPheLysLeuGlyGI 120
Db 16992 GGGCGGCTGTGACGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 17051

```

```

QY 120 yAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsn----- 136
Db 17052 GACCTGTCCGCGCGCGC---GAGGACACAGTCCCTCCGACCCGAGGACGTACGC 17108
QY 137 -----AsnAspAsnI 140
Db 17109 CCGCCGTGTGCAGGGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 17168
QY 140 eLeuTyrPheGlyAlaProSerGlyLysGly-----LeuTyr----- 152
Db 17169 GCTCGTCACCGCGGCGCACCGCGGCGGTGCGTCCGACGTCGCGCGGTGCTGCCGCCA 17228
QY 153 -----ArgSerThrAspSerGlyAlaThrTrpSe 162
Db 17229 GGGCACCCCGTGTGCTGTGCGCACCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 17288
QY 162 rGlnMetThrAsnPheProAspValGlyThrTyrIleAla-AsnPro-----ThrAspT 180
Db 17289 GCTACTCACCGAACTCCCGACCTGGGCGACCCCGGCGCACCGTCCCGCTCGGACGTCA 17348
QY 180 hrThrGlyTyrGlnSerAspIleGlnGlyValValTyrValAlaPheAspLysSerSers 200
Db 17349 CGACCGGAGCAGCTCC-----GTGCCCTCTCCGCGACCGTCCGACGACGAGA 17396
QY 200 erSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProv 220
Db 17397 CCGCTGTCCGCGGTGTTCCAGTCGCGCGGCGACGTCGACGACGCGACCGTCCGACGCT 17456
QY 220 alPheTyrSerArgAspGly-----AlaThrTrpG 231
Db 17457 CACCGGTGACCGCATCGAACCGGCGCAACCGGCGGAGGTGCTCGGTGCGCGCACTGCA 17516
QY 231 lAlaValProGlyAlaProGlyPheIleProHisLysGlyValPheAspProVala 251
Db 17517 CGAGCTACCGGCGACCGCGACCTCG----- 17542
QY 251 snHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyA 271
Db 17543 -----ACGCGTTCGTGCTCTCTCTCCACCGCGGCGGCGGCGCGCGCGGCGG- 17593
QY 271 spValTyrLysPheSerValThrSerGlyThrTyrThrArgIleSerProVal----- 288
Db 17594 -----TCGCGCGGTACGTCGCGGCGCAACGCTACTCTCGACGG 17630
QY 289 -----ProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspA 307
Db 17631 TCTCGCCACGACGCGACGCGAGG-----GACT 17660
QY 307 rGlnHisProAsnThrIleMetValAlaThrGlnIleSerTyrTrpProAspThrIleI 327
Db 17661 CCGCGCCACCTCGGTGCGGTGCGGTACCTGGCGCGGCGAGCGGATGGCGAGGCTC--- 17716
QY 327 lePheArgSerThrAspGlyClyAla-----ThrTrp----- 337
Db 17717 -----CGGTGCGCGACCGGTTCGCGCGGCGACGGGTTCATGGAGATGCACCCCGACCGC 17771
QY 338 --ThrArgIle-----TyrAspTrp----- 343
Db 17772 CGTCGGGTGCTCCGGTGGCGACTGGTGGAGGTGAGTAGTACCCCGATCGTCTCGTACAT 17831
QY 344 -----ThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluP 361
Db 17832 CAGGTGGACCGGTTCCTCTCGGTACACCGCGGCGGCGGCGGCGGCGGCGGCTCTTCGACAC 17891
QY 361 roTyrLeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGlyTyrPM 381
Db 17892 CC---TCGACGAGGCGGTGCGGCGGCGC---CGGTGCGCGACGCGGCGGCGGCGGCTGC 17945
QY 381 et----- 381
Db 17946 G-CGCTGCGCGGCTCGCGTCCGCGGAAACGCGAGAGCGCGGTCTCTCGACCTGTGTACGA 18004
QY 382 -----AspGluAlaM 385

```

Db 18005 CGCAGCGGCTCCGCTCGGCCACGCTCGCGCAGCAGGTGCCGTCGACGGGCT 18064  
QY 385 etAla-----IleAspProPheAsnSer-----AspArgMetLeuTyrG 398  
Db 18065 TCGCCGAACCTCGCGCTCGACTCGCTGTCGGCCCTGGAACTGGCAACCGGTCGACCACTG 18124  
QY 398 lYThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAsp----- 413  
Db 18125 CGACCGGGTCCGGTGGCCACGACGCGGTCTTCACACCGGAGGTCGACGACCTGG 18184  
QY 414 -----serGlyGlyGlnI 418  
Db 18185 CGGACACCTGCGCCGCAACTGGCGCGGATCGCGGCGGAGCGGCCCGCGGCGAG- 18243  
QY 418 leHisIleAlaProMetValLysGlyLeuGluGluThrAlaVal----- 432  
Db 18244 -----GCCCCGAGTGGCCCGACGACGAGCGGATCCCATCGTCGGGATGGCCT 18295  
QY 433 -----AsnAspLeuIleSerProSerGlyAlaProLeuIleSerAlaLeuG 449  
Db 18296 GCCGGTCCGGGGAGTGGACTCACCGAGCAGCTGGGAGTTCGTCCTCCGGC 18355  
QY 449 lYAspLeuGlyGlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrS 469  
Db 18356 GGGACACCGCTCGCGCGGCACCGCGGACCGGAGTGGGATCCGCGGAGTTCGATGCT 18415  
QY 469 exProValPheThrGlyThrGlySerValAspTyrAlaGluLeuAsnProSerIleIleV 489  
Db 18416 CGGAC-----ACGACGGGACCCCGTACCGCTTCGGCAATTCATGCCCGG- 18462  
QY 489 aLArgAlaGlySerPheAspPro-----Sers 498  
Db 18463 -----GCGGCGAGTTCAGCGGGGCTTCTCGGATCTCGCGCGTGGAGGCTGGCGA 18517  
QY 498 exGlnProAsnAspArgHis----- 504  
Db 18518 TGGATCCGACGAGCGGACGCGCTGGAGACCACTGGGAGCGCTGGAGAACCGCGTA 18577  
QY 505 -----ValAlaPheSerThrAspG 511  
Db 18578 TCCGGCCCGAGTCGTTGGCGGTACGGACACACCGGTGTCTTCGTGGGATGTCCTCATCAG 18637  
QY 511 lY-----GlyLysAsnTrpPheGlnGlySerGluProGlyGlyValThrGlyG 528  
Db 18638 GTTACGCCACCGCGCCCGAAGCCGAGGAGGTGACGCGCTACCTGTTGACAGGCA 18697  
QY 528 lYThrValAlaAlaSerAlaAsp-GlySerArgPheValTrpAlaProGlyAspProGly 547  
Db 18698 ACACCGGAGCTCGCTCCGCTCGGTGCGATCGCTGTTGGGT- -GGAGGGGCGGC 18755  
QY 548 GlnProValValTyrAlaVal-----GlyPhe 556  
Db 18756 GATCACTGTGGACACGGGTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 18815  
QY 557 -----GlyAsnSerTrpAlaAlaSer----- 563  
Db 18816 GTTGGTCTCGGGACTGTGCTGCGGTGCGGTGGGTGCTGCTGCTGCTGCTGCTGCTGCT 18875  
QY 564 -----GlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLys 580  
Db 18876 GGAGGTGTCAGGAGTTCCTCCGCGAGGCGCGGTGGCTCGGACGG-----CAG 18926  
QY 581 ThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPhe 600  
Db 18927 GTGCAAGCCCTTCTCGGACGAGGCGGACGCGCTTCGCTCT- -GGGAGGCGGTTCGCGCTT 18983  
QY 601 GlnProValAlaAla-----GlyLeuProSerSerGlyAlaValGlyValMetPheHis 618  
Db 18984 CTTGCTGTTGACGCGGTGTCGCTGCGGTGCGGAGGCGGCTGCGGTGTTGGGTGTGCT 19043  
QY 619 AlaValProGlyLysGlu-----GlyAspLeuTrpLeuAlaAlaSerSerGlyLeu 635

Db 19044 GGTGGTTCGGCGGTGAATCAGATCGGGCGAGTAATGGTGGCGGCGCGCTCGGG- 19100  
QY 636 TyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaVal 655  
Db 19101 -----GGTGGCGCAGCAGCGGTGATTCGGCGGCGGTGGGTCTGTCGGG 19145  
QY 656 AsnValGlyPheGlyLysSerAlaProGlySerSerTyr-----ProAlaValPheVal 673  
Db 19146 TGTGTCGGGTGGGATGTCGGTGTGGTGGAGGCGCATCGGACGGGACCGCGTTGGGGA 19205  
QY 674 ValGlyThrIleGlyGlyValThrGlyAlaTyrArgSerAspAspCysGlyThrTrp 693  
Db 19206 TCGGTGGATTCGGGCGTGTGGG----- 19232  
QY 694 ValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAsp 713  
Db 19233 -----GACGTATCGGTGGTGGGTGGGT----- 19259  
QY 714 HisAlaAsnLeuArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAsp 733  
Db 19260 -----GGTCCGCTGCTGCTGGTTCGGTGAAGCGCA 19292  
QY 734 IleGlyGlyAlaProSerGly 740  
Db 19293 TGTGGTTCATGTCACGCGGC 19313  
RESULT 15  
US-09-252-991A-5730  
; Sequence 5730, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 5730  
; LENGTH: 4647  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-5730  
Alignment Scores:  
Pred. No.: 0.000254 Length: 4647  
Score: 159.50 Matches: 198  
Percent Similarity: 30.51% Conservative: 90  
Best Local Similarity: 20.97% Mismatches: 284  
Query Match: 3.95% Indels: 372  
DB: 4 Gaps: 57  
US-09-917-376-3 (1-740) x US-09-252-991A-5730 (1-4647)

QY 43 ArgTrpAspAlaAlaAsnGlyArgTrpIle-----ProLeuLeuAsp-----TrpVal 58  
Db 861 CAATGGAGAGTGGCTTGGCACGTGGTCTCTGATCCCTTCGACAGCAGGAGTGGTACT 920  
QY 59 GlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAla-AlaAspProIleAsnTh 78  
Db 921 GGCTGG- -CGATGCGCTTTCATCGAGTTGCG- -GCCACGACGCGGTGTCGCGCA 971  
QY 78 AsnLysValTrpAlaAlaValGlyMetTyr-----ThrAsnSerTrpAsp----- 93  
Db 972 GGGACCAACTGGGCATTTGGCCGAGATGTCGCCGCCGCTAGCCAGTCTGTCGTCATCGG 1031  
QY 94 -----ProAsnAspG1 97  
Db 1032 CCTGGTCCGCGGTGCCGATCTCGCGGTGCCGACACCGTCTGGTGGTCCGACTCCTC 1091

QY 97 yAlaIleLeuArg----- 101  
 Db 1092 GCGCAGCTGAACCTGGCGGATACGCACTATCAGGCGAAGATCGAACAGAGCTCGGGAGG 1151  
 QY 102 -----SerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPhePheLeuGlu 119  
 Db 1152 GCTGGTGTTCACCGATCAGCTACCGACTCGGGGATCACCCTG----- 1194  
 QY 119 yGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAs 139  
 Db 1195 -----GGAACCTCCGTCGACGAGCAATGAATG 1223  
 QY 139 nIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTyrArgSerThrAspSerGlyAl 159  
 Db 1224 GATC-----TGCGGCTGGGCGGTATTGTGCGGAGCTCCCGCTG 1265  
 QY 159 aThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAs 179  
 Db 1266 GACCTGGGCGCC-----GGCAATTACCTGGGCATGCTGCTGG 1304  
 QY 179 pThrThrGlyTyrGlnSerAspIleGlnGlyValValTyrValAlaPheAspLysSer-- 198  
 Db 1305 TACGGCGATCGCGGCGGCACTGCG-----TGTGCACTGTGATCCGTCCT 1355  
 QY 199 -----SerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAs 217  
 Db 1356 GTGTATCGAGAACTCTCGGCACAAACC-----GTGTCACACCGCA 1394  
 QY 217 nAsnProValPhe----- 221  
 Db 1395 GAACCATGTTGTTTCAGTGTGTGCGCACCGCACTGGCGATCTTGACCTGGCCAGTGGCCG 1454  
 QY 222 -----TrpSerArgAspGly-----GlyAlaThrTrpGlnAlaValProG 235  
 Db 1455 TAACCTTACGCAATGTCGCGTACGCGCTCTATACCGCGGAACGAGCGCCCGCATGT 1514  
 QY 235 yAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHisValLeu-- 254  
 Db 1515 CGCG--ACGGGATTCAATCAGCCTCGCGTCTGTTCAAT-----GGGTCTGCTACTGG 1565  
 QY 255 -----TyrIleAlaThrSer-----As 260  
 Db 1566 CGCGCGTGGCGGATTCAGAACTGTCTCAACAGCAGCTACCGACCTGGTATCCGGA 1625  
 QY 260 nThrGlyGlyProTyrAsp-----GlySerSer-----GlyAspValTrpLysPh 275  
 Db 1626 ACATGCGCGCAACTCGATATCGCGTGGCGGTGATGTGTTGGCGATCAATGGGCGCA 1685  
 QY 275 eSerValThrSerGlyThrTrpThrArgIleSerProValProSerThrAspThrAlaAs 295  
 Db 1686 GAAACTGACAGTAGC-----GATCCCATTCGCCCCCTGGCCCCCAGCGCTGCAGTCCG 1739  
 QY 295 nAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetVa 315  
 Db 1740 CAACCTGGCTCTGGCGCCAGGCG-----AGCGCGCATAGGGAAGAGTCCCGACG----- 1788  
 QY 315 lAlaThrGlnIleSerTrpThrProAspThrIleIlePhe--ArgSerThrAspGlyGl 334  
 Db 1789 -----GCTGTGGGTGCAATTTCGCGACGTATGTGCGTGGCGCGGAGGAGA 1835  
 QY 334 yAlaThrTrpThrArgIleTrpAspTrpThr----- 344  
 Db 1836 TGCTCCCTAT-----CTGGTGGCTTTCACCGGTTTCGGCACTCTGGGAGCGGCAATCT 1889  
 QY 345 -----SerTyrPr 347  
 Db 1890 CAGCATGCGAACCGGAGGTGATCGCGCAATATCGCGCCACGCGGAGATGGCTCGATACC 1949  
 QY 347 oAsn-----ArgSerLeuArgTyrValLeuAspIleSerAlaGluPr 361  
 Db 1950 TTCGTCCGGTAACCTTAATCCCGGCGACCGAGGACTGGTCTTGGCGGTTCGCGGCACTGG 2009

QY 361 oTrpLeuThr-----Ph 365  
 Db 2010 CCGGCTCACCTCCGACGCGCGCTGCAACTTGGGCGCGGTGACCTGAATGTGAGAA 2069  
 QY 365 eGlyValGlnProAsnProProValProSerProLysLeuGlyTyrMetAspGluAlaMe 385  
 Db 2070 TGGCGGGAGGTGAACCCG-----AGCGGGAGGCGCG 2102  
 QY 385 tAlaIleAspProPheAsnSerAspArgMet-----LeuTyrGlyThrGly----- 400  
 Db 2103 GGCTACCAGACCTACTCGAGTTCGGGGTTTGACCGGTCTTTACAGCGGCGGACCATTCA 2162  
 QY 401 -----AlaThrLeuTyrAlaThrAsnAs 408  
 Db 2163 CGATCTGCGAGGGCGCCCTCATCAATCTGCGCGGTTCCGCGAGCTTGATC----- 2211  
 QY 408 pLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaPro-----MetVa 424  
 Db 2212 -----AGTGGTGGCTGGGAGGATCGATCCTCGCTAGCACACTCTACT 2255  
 QY 424 lLysGlyLeuGluGluThrAlaValAsnAspLeuIleSerPro-----Pr 439  
 Db 2256 GCGGACCCCGCCAGAGTGGCGCAGCGGATGCGTTAGCCCGACGCTGGCTAGCTCCAC 2315  
 QY 439 oSerGlyAlaProLeuIleSer-----AlaLeuGlyAsp-- 450  
 Db 2316 GGTGGGCTCACCTGCTGGCGGAGATACTGGCATGCGTCTGGAACCCGCTGGCGACCT 2375  
 QY 451 -----LeuGlyGlyPheThrHisAlaAspValThrAlaValProSerThrIle--PheTh 468  
 Db 2376 GGTCTGGTGGGTGAGTACTACCCCGTGGGTGGGTGGGATACCGTTGGCTTTCAC 2435  
 QY 468 rSerPro-----ValPheThrTh 474  
 Db 2436 GGCTCCGAGTGGCAGCTCTACCAAGGGGGTGGGATCGTTGGTTCAGTCTCTGACTGC 2495  
 QY 474 rGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPh 494  
 Db 2496 ACACACTCGATCGAT-----CTGTTTGGCGGGCGGCGCAACCT 2534  
 QY 494 eAspProSerSerGln----- 499  
 Db 2535 GACGCCAGCACCAATTTGGTCGAGGCAACCAACGCTATTCCCATGGCAGGCGCAACCT 2594  
 QY 500 -----ProAsnAspArgHisValAlaPheSerThrAsp----- 510  
 Db 2595 GTCGCCCTAGCGTGGCGATCATCTATCCATCGATAGTAGTCGCGGGTGGCCCGGAAGG 2654  
 QY 511 -----GlyGlyLysAsnTrpPheGlnGlySerGluProGlyGlyValThrTh 526  
 Db 2655 AAGTATCTATCTCGTCTCTCCAGCGGTTACATGGGAGCGGTGAGCCTGAATGTCTCGAC 2714  
 QY 526 r-----GlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPr 543  
 Db 2715 CACTCCCTACTCTTGTGTACTGGCACCGCTCGCTCAACGGTGAACCTGGAGTTCTCGCT-- 2772  
 QY 543 oGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSe 563  
 Db 2773 -GGAGACTCC-----ATTATGCGCGTGGCTAT-----TCGTCACAGCTC 2813  
 QY 563 rGlnGlyValProAlaAsnAlaGlnIle-----ArgSerAs 575  
 Db 2814 AGGTGCGCATCTCTGGAACCTGCCATCGATCTGGACGCGCGCATTTGCGGCTACTCCGA 2873  
 QY 575 pArg-----ValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSe 593  
 Db 2874 TGCGCGCTGTCTCAACCG-----ATAGCGGGCAACCGC-----AG 2909  
 QY 593 rThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaVa 613  
 Db 2910 TCCGACACGGGAAT-----CCGCGCGTCAATTGGCGGATTCCT----- 2946  
 QY 613 lGlyValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSe 633



```
Db 2947 -----CTGTCTACTCGGCCCGATAGCGCTGGAGCCTGGCGCGGATCTGCAACC 2999
QY 633 rGlyLeuTyrHisSerThrAsnGlyGlySerTrpSerAlaIleThrGlyValSerSe 653
Db 3000 GCGCGGTTTCTATGCTTACACAGAGAT-----ATCGTCGGGCTCAATAG 3044
QY 653 rAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheVa 673
Db 3045 CGGTGCGCAGATTCTGTTTCGGCGAGCAGCG----- 3075
QY 673 lValGlyThrIleGlyGlyValThrGlyAlaTyrArgSerAspCysGlyThrThrTr 693
Db 3076 -----GGGAACCGCGCGCGCAACCTGGTATGAAGTGGGGTCCCGTCTG 3122
QY 693 pValLeuIleAsnAspAspGlnHisGlnTyrGlyAsn---TrpGlyGlnAlaIle----- 710
Db 3123 GATGCGCGCGCGCGACATCGTTGCCCTCCGGAACCCCATTTGGACACGCGTATCAGCGC 3182
QY 711 -----ThrGlyAsp-----HisAlaAsnLe 717
Db 3183 TCCAGTCAGATCAGTACGATGCTCGTTACCGCAACCTCTTCGTCACGACGACCC 3242
QY 717 uArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyAl 737
Db 3243 CAACGACCTCTCGCTGGTCCAGGCGCGCGCGACATCCTCTACGGCAACTTCAACGTCGC 3302
QY 737 aProSerGly 740
Db 3303 CGGCCCCGGT 3312
```

Search completed: May 12, 2004, 15:15:19  
Job time : 12944.8 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 12, 2004, 09:26:52 ; Search time 1136.23 Seconds  
(without alignments)  
2950.286 Million cell updates/sec

Title: US-09-917-376-3

Perfect score: 4036

Sequence: 1 ATTQPTYWGNVAIGGGGFVD.....YIGTNGRGIYVIGIGAPSG 740

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2941586 seqs, 2264995651 residues

Total number of hits satisfying chosen parameters: 5883172

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DBV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US09917376/runat\_11052004\_114537\_28336/app\_query.fasta\_1.1429  
-DB=Published Applications NA -QFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCLALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09917376@cgn\_1\_1\_809@runat\_11052004\_114537\_28336  
-NCPU=6 -ICPU=3 -NO MMAP -LARGETIMEOUT=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=120 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq:\*  
17: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
18: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
19: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

ALIGNMENTS

RESULT 1

US-09-917-376-2  
; Sequence 2, Application US/09917376  
; Publication No. US20040038334A1

; GENERAL INFORMATION:

; APPLICANT: DING, SHI-YOU

; APPLICANT: ADNEY, WILLIAM S.

; APPLICANT: VINZANT, TODD B.

; APPLICANT: HIMMEL, MICHAEL E.

; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS

; TITLE OF INVENTION: CELLULOXYTICUS

; FILE REFERENCE: 40197.4US01

; CURRENT APPLICATION NUMBER: US/09/917,376

; CURRENT FILING DATE: 2001-07-28

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 2869

; TYPE: DNA

; ORGANISM: Acidothermus cellulolyticus

; FEATURE:

; NAME/KEY: modified\_base

; LOCATION: (2869)

1	4036	100.0	2869	13	US-09-917-376-2	Sequence 2, Appli
2	4036	100.0	2869	15	US-10-155-400-2	Sequence 2, Appli
3	2429.5	60.2	2646	15	US-10-156-761-1845	Sequence 1845, Ap
4	2429.5	60.2	9025608	15	US-10-156-761-1	Sequence 1, Appli
5	1579	39.1	2517	15	US-10-026-994-4	Sequence 4, Appli
6	1579	39.1	2710	15	US-10-026-994-1	Sequence 1, Appli
7	1405	34.8	2217	15	US-10-156-761-2561	Sequence 2561, Ap
8	1405	34.8	9025608	15	US-10-156-761-1	Sequence 1, Appli
9	1139	28.2	3668	10	US-09-927-827-21	Sequence 21, Appli
10	1117	27.7	2367	13	US-10-395-241-13	Sequence 11, Appli
11	1117	27.7	2646	13	US-10-395-241-11	Sequence 11, Appli
12	1114	27.6	2481	13	US-10-395-241-17	Sequence 17, Appli
13	878.5	21.8	5698	13	US-10-420-191-1	Sequence 1, Appli
14	238	5.9	7407	13	US-10-282-122A-30151	Sequence 30151, A
15	238	5.9	7407	13	US-10-246-330-3	Sequence 3, Appli
16	199.5	4.9	7305	13	US-10-282-122A-11269	Sequence 11269, A
17	195	4.8	5295	13	US-10-282-122A-11871	Sequence 11871, A
18	192	4.8	9903	13	US-10-282-122A-28185	Sequence 28185, A
19	188.5	4.7	4399	16	US-10-369-493-37642	Sequence 37642, A
20	184	4.6	3324	13	US-10-282-122A-26288	Sequence 26288, A
21	182	4.5	69350	12	US-10-041-018-379	Sequence 379, App
22	180	4.5	2355	16	US-10-369-493-37792	Sequence 37792, A
23	179.5	4.4	3222	13	US-10-282-122A-14432	Sequence 14432, A
24	176	4.4	32367	15	US-10-158-160A-14	Sequence 14, Appli
25	170.5	4.2	7191	13	US-10-282-122A-35048	Sequence 35048, A
26	170.5	4.2	25165	15	US-10-114-170-39	Sequence 39, Appli
27	170	4.2	4323	13	US-10-282-122A-8383	Sequence 8383, Ap
28	169.5	4.2	3360	15	US-10-245-802-11	Sequence 11, Appli
29	169	4.2	69350	12	US-10-041-018-379	Sequence 379, App
30	168	4.2	5451	15	US-10-101-510-597	Sequence 597, App
31	167.5	4.2	2110	13	US-10-027-632-97984	Sequence 97984, A
32	167.5	4.2	2110	16	US-10-027-632-97984	Sequence 97984, A
33	167	4.1	1839	15	US-10-369-493-34868	Sequence 34868, A
34	166	4.1	3879	13	US-10-282-122A-8717	Sequence 8717, Ap
35	165.5	4.1	4383	15	US-10-245-802-21	Sequence 21, Appli
36	165	4.1	6615	13	US-10-282-122A-28180	Sequence 28180, A
37	164.5	4.1	3237	12	US-10-152-319A-1871	Sequence 1871, Ap
38	163	4.0	5145	9	US-09-925-299-206	Sequence 206, App
39	163	4.0	5145	10	US-09-925-299-206	Sequence 206, App
40	161.5	4.0	4812	15	US-10-084-817-4	Sequence 4, Appli
41	161.5	4.0	5432	14	US-10-044-090-22	Sequence 22, Appli
42	161	4.0	2712	15	US-10-156-761-3375	Sequence 3375, Ap
43	159.5	4.0	3129	16	US-10-369-493-42336	Sequence 42336, A
44	159.5	4.0	10419	14	US-10-027-806-3	Sequence 3, Appli
45	159.5	4.0	10419	14	US-10-034-623-3	Sequence 3, Appli

OTHER INFORMATION: a, c, t, g, other or unknown  
US-09-917-376-2

Alignment Scores:

Pred. No.: 0 Length: 2869  
Score: 4036.00 Matches: 740  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 13 Gaps: 0

US-09-917-376-3 (1-740) x US-09-917-376-2 (1-2869)

QY 1 AlaThrThrGlnProTyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAsp 20  
DB 139 GCGAGACTCAGCCGTACACCTGGAGCAACGTTGGCGATCGGGGGCGGGCTTTGTCCGAC 198  
QY 21 GlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGly 40  
DB 199 GGGATCGCTCTCAATGAAGGTGCACCGGAATTCGTACGTGCGGACGACATCCGGGGG 258  
QY 41 MetTyrArgTTPAspAlaAlaAsnGlyArgTTPilleProLeuLeuAspTTPValGlyTTP 60  
DB 259 ATGTATCGATGGATGCGCCCAACGGGGTGGATCCCTCTTCYGGATTGGTGGGATGG 318  
QY 61 AsnAsnTTPGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 80  
DB 319 AACAAATGGGGGTACAACGGGCTGTCAGCATTCGGCGAGACCCGATCAATACTAACAG 378  
QY 81 ValTTPAlaAlaValGlyMetTyrThrAsnSerTTPAspProAsnAspGlyAlaIleLeu 100  
DB 379 GTATGGGGCGCGCTCGGAATGTACACCAACAGCTGGGACCCCAACAGCGGACGATTCCTC 438  
QY 101 ArgSerSerAspGlnGlyAlaThrTTPGlnIleThrProLeuProPheLysLeuGlyGly 120  
DB 439 CGCTCGTCTGATCAGGGGCGCAACGTGGCAATAACGCCCTTCGCCCTTCAAGCTTGGCGGC 498  
QY 121 AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle 140  
DB 499 AACATGCCCGCGGTGAATGGCGAGCGGCTTGGGTGGATCCAAACAATGACAAACATT 558  
QY 141 LeuTyrPheGlyAlaProSerGlyLysGlyLeuTTPArgSerThrAspSerGlyAlaThr 160  
DB 559 CTGTATTTTCGGCGCCCGCGAGCGCAAGGGCTCTGGAGAAGCACAGATTCGGCGCGGACC 618  
QY 161 TTPSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThr 180  
DB 619 TGGTCCCAGATGACGAATTTCCGAGCTAGCGCATACATTGCAAAATCCCACTGACACG 678  
QY 181 ThrGlyTyrGlnSerAspIleGlnGlyValValTTPValAlaPheAspLysSerSerSer 200  
DB 679 ACCGGCTATCAGAGCGATATTCAAGCGCTGCTTGGGTGGCTTTCCGACAAAGTCTTCGTCA 738  
QY 201 SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProVal 220  
DB 739 TCGCTCGGGCAAGCGAGTAAGACCAATTTTGTGGCGCTGGCGGATCCCAATAATCCGGTTC 798  
QY 221 PheTTPSerArgAspGlyGlyAlaThrTTPGlnAlaValProGlyAlaProThrGlyPhe 240  
DB 799 TTTCTGGACAGAGACGGCGCGCGAGCGGTGGAGGGGTGGCGGGTGGCGCGCGCGCTTC 858  
QY 241 IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsn 260  
DB 859 ATCCGCGACAAAGGGCTCTTTGACCCGCTCAACACAGTGTCTATATTGCCACACGACAT 918  
QY 261 ThrGlyGlyProTyrAspGlySerSerGlyAspValTTPLysPheSerValThrSerGly 280  
DB 919 ACGGGTGGTCCGTATGACGGGAGCTCCGCGCAGCTCTGGAATTCCTGGTGTACCTCCGGG 978  
QY 281 ThrTTPThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyr 300  
DB 979 ACATGGAGCGGAATCAGCCCGGTACCTTCGACGGACACGGCCCAACGACTACTTTGGTTAC 1038

QY 301 SerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSer 320  
DB 1039 AGCGGCTCACTATCGACGCCAGCACCCGAAACAGATATATGTGGCAACACGATATCG 1098  
QY 321 TTPTTPProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTTPThrArgIle 340  
DB 1099 TGTGGCCCGGACACCATATCTTTCGGAGACCCGACGGCGGTGCGAGTGGACGCGGATC 1158  
QY 341 TTPAspTTPThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGlu 360  
DB 1159 TGGGATTGGACGAGTTATCCCAATCGAAGCTTCGGATATGTGCTTGACATTTCCGGCGAG 1218  
QY 361 ProTTPLeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGlyTTP 380  
DB 1219 CCTTGGCTGACCTTCGGCGTACAGCGAATCCTCCCGTACCCAGTCCGAAAGTCCGGCTGG 1278  
QY 381 MetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetIleuTyrGlyThrGly 400  
DB 1279 ATGGATGAAGCGATGGCAATCGATCCGTTCACCTCTGATCGGATGCTCTACGGAACAGGC 1338  
QY 401 AlaThrLeuTyrAlaThrAsnAspLeuThrLysTTPAspSerGlyGlyGlnIleHisIle 420  
DB 1339 GCGACGTTGTACGCAACAATGATCTCAGAACTGGGACTCCGGCGGCGAGATTCATATC 1398  
QY 421 AlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIleSerProProSer 440  
DB 1399 GCGCCGATGGTCAAAAGGATTGGAGGAGACGGCGGTAAACGATCTCATCAGCCCGCTCT 1458  
QY 441 GlyAlaProLeuIleSerAlaIleuGlyAspLeuGlyGlyPheThrHisAlaAspValThr 460  
DB 1459 GCGGCCCTCATCAGCGCTCTCGGAGACCTCGCGGGCTTCCACCACGCGCGATTTACT 1518  
QY 461 AlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyr 480  
DB 1519 GCGGTGCCATCGACGATCTTTCAGTCCCGGTGTTCCAGCCCGGACCGAGCTGCATAT 1578  
QY 481 AlaGluLeuAsnProSerIleValArgAlaGlySerPheAspProSerSerGlnPro 500  
DB 1579 GCGGAATTAATCCGTCGATCATCGTTCGCGCTGGAAGTTTCGATCCATCAGGCAACCG 1638  
QY 501 AsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTTPPheGlnGlySerGlu 520  
DB 1639 AACGACAGCACGTCGCTTCTCGACAGACGGCGGCAAGAACTGGTTCCAAGGCGAGCAA 1698  
QY 521 ProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheVal 540  
DB 1699 CTTGGCGGGGTGACGACGGCGGCGCCGTCGCGCATCGGCCGACGCTCTCGTTTCGTC 1758  
QY 541 TTPAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTTP 560  
DB 1759 TGGGCTCCCGCGATCCCGGTGAGCTGTGTTGTTACGAGTCGGAATTTGGCAATCCTCTGG 1818  
QY 561 AlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLys 580  
DB 1819 GCTGCTTCGCAAGGTGTTCCCGCAATGCCAGATCCGCTCAGACCGGGTGAATCCAAAG 1878  
QY 581 ThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPhe 600  
DB 1879 ACTTTTATGCTCCCTATCCCAATGAACTTCTATCGAAGACGACGCGCGCGGTGACATTC 1938  
QY 601 GlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaVal 620  
DB 1939 CAACCGTCCGCGCGGTCTTCCGAGCAGCGGTGCGCTCGGTGTCATGTTCCACGCGGTG 1998  
QY 621 ProGlyLysGlyAspLeuTTPLeuAlaLaserSerGlyLeuTyrHisSerThrAsn 640  
DB 1999 CCTGGAAGAAAGGCGATCTGTGGCTCGCTGTCATCGAGCGGGCTTTACCACTCAACCAAT 2058  
QY 641 GlyGlySerSerTTPSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGly 660  
DB 2059 GCGCGCAGCAGTGTGTGCAATCAACCGCGGTATCTCCGCGGTGAACGGGATTTGGT 2118  
QY 661 LysSerAlaProGlySerSerTyrProAlaValPheValGlyThrIleGlyVal 680

```
Db 2119 AAGTCTGGCCCGGTCGTATACCCAGCCGCTCTTTGTCGCGCACATCGGAGCGGT 2178
Qy 681 ThrGlyAlaTyrArgSerAspCysGlyThrThrTTPValLeuLeuAsnAspGln 700
Db 2179 ACGGGGGGTACCGCTCCGACGACTGTGGACGACCTGGGTACTGTATCATGTATGACCA 2238
Qy 701 HisGlnTyrGlyAsnTTPGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgVal 720
Db 2239 CACCAATACGGAATTTGGGACAGCAATACCGGTGACCAACGCAATTTTACGGCGGTG 2298
Qy 721 TyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyAlaProSerGly 740
Db 2299 TACATAGCACGAACCGCGGTGAATTTATACGGGACATTTGGTGTGCGCCGTCGGA 2358

RESULT 2
US-10-155-400-2
; Sequence 2, Application US/10155400
; Publication NO. US2003010988A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOYTICUS
; FILE REFERENCE: NREL 01-36A
; CURRENT APPLICATION NUMBER: US/10/155,400
; CURRENT FILING DATE: 2002-10-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2869
; TYPE: DNA
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (2869)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-10-155-400-2

Alignment Scores:
Pred. No.: 0 Length: 2869
Score: 4036.00 Matches: 740
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 15 Gaps: 0

US-09-917-376-3 (1-740) x US-10-155-400-2 (1-2869)
Qy 1 AlaThrThrGlnProTyrThrTTPSerAsnValAlaIleGlyGlyGlyPheValAsp 20
Db 139 GCGACGACTCAGCCGTACACCTGGAGCAACCTGGCGATCGGGGCGCGGCTTTGTCGAC 198
Qy 21 GlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValAcqThrAspIleGly 40
Db 199 GGGATCGTCTTCAATGAAGGTGCACCGGAATTCGTACGTGCGGACGACATCGGGGG 258
Qy 41 MetTyrArgTTPAspAlaAlaAsnGlyArgTTPileProLeuLeuAspTTPValGlyTTP 60
Db 259 ATGTATCATGATGGATGCCGCCACCGGGCGGTGGATCCCTCTCTGGATTTGGGTGGATGG 318
Qy 61 AsnAsnTTPGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 80
Db 319 AACAAATGGGGGTACAACGGCGTGTACAGCTTGGCGGACACCCGATCAATACTAACAA 378
Qy 81 ValTTPAlaAlaValGlyMetTyrThrAsnSerTTPAspProAsnAspGlyAlaIleLeu 100
Db 379 GTATGGCGCGCGTTCGGAATGTACACCAACAGCTGGGACCCAAACGACGGAGCGATTCTC 438
Qy 101 ArgSerSerAspGlnGlyAlaThrTTPGlnIleThrProLeuProPheLysLeuGly 120
```

```
Db 439 CGTCTGTGATCAGGGCGCAACGTGGCAATAACGCCCCCTGCGCTTCAAGCTTGGCGGC 498
Qy 121 AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle 140
Db 499 AACATGCCCGGGCGTGAATGGCGAGCGCTTGGGTGGATCCAAACAATGACAACATT 558
Qy 141 LeuTyrPheGlyAlaProSerGlyLysGlyLeuTTPArgSerThrAspSerGlyAlaThr 160
Db 559 CTGTATTTTCCGCGCCCGGAGCGCAAGGGCTCTGGAGAAGCACGATTTCCCGCGGACC 618
Qy 161 TTPSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThr 180
Db 619 TGGTCCCAGATGACCAACTTTCCGACGTAGGACGTACTATTCGAATCCCACGTGACAG 678
Qy 181 ThrGlyTyrGlnSerAspIleGlnGlyValValTTPValAlaPheAspLysSerSer 200
Db 679 ACCGGCTATCAGAGCGATATTCAGGGCTGCTTGGTGGCTTTCGACAAGTCTTCGTCA 738
Qy 201 SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnProVal 220
Db 739 TCGCTCGGCGCAAGCGAGTAAGACCAATTTTGTGGCGTGGCGATCCCAATAATCCGGTC 798
Qy 221 PheTTPSerArgAspGlyGlyAlaThrTTPGlnAlaValProGlyAlaProThrGlyPhe 240
Db 799 TTCTGGACGAGACGCGCGCGCGACGTGGCAGGCGTCCGCGTGGCGGACCGGCTTC 858
Qy 241 IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsn 260
Db 859 ATCCCGCACCAAGGGCGTCTTTGACCGCGTCAACACGCTCTATATTGCCACAGCAAT 918
Qy 261 ThrGlyGlyProTyrAspGlySerGlyAspValTTPLysPheSerValThrSerGly 280
Db 919 ACGGTGTGTTCGTATGACGGGAGCTCCGCGCAGCTCTGGAAATTTCTCGGTGACTCCGGG 978
Qy 281 ThrTTPThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyr 300
Db 979 ACATGGACGCAATCAGCCCGGTACTCTTCGACGACACGCGCAACACACTACTTTGGTAC 1038
Qy 301 SerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSer 320
Db 1039 ACGGCGCTCATTATCGACCGCAGCACCCGAAACACGATTAATGGTGGCAACCCAGATATCG 1098
Qy 321 TTPTTPProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTTPThrArgIle 340
Db 1099 TGGTGGCGCGGACACCATATCTTTCCGACACCGACGGCGGTCCGACGTGGACGGGATC 1158
Qy 341 TTPAspTTPThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGlu 360
Db 1159 TGGGATTGGACGAGTTATCCCAATCGAGCTTGGCATATGTGCTTGACATTTCCGGCGAG 1218
Qy 361 ProTTPLeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGlyTTP 380
Db 1219 CTTTGGCTGACCTTCGGCGTACAGCGAATCTCTCCGTACCCAGTCCGAGCTCGGCTGG 1278
Qy 381 MetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGly 400
Db 1279 ATGGATGAAGCGATGGCAATCGATCCGTCAACTCTGATCGGATGCTCTACGGACAGGC 1338
Qy 401 AlaThrLeuTyrAlaThrAsnAspLeuThrLysTTPAspSerGlyGlyGlnIleHisIle 420
Db 1339 GCGACGTTGTACGCAACAAATGATCTACGAAGTGGGACTCCGGCGCGCCAGATTTCATATC 1398
Qy 421 AlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIleSerProSer 440
Db 1399 GCGCGCATGGTCAAAAGGATTGGAGGAGACGGCGGTAAACGATCTCATCAGCCCGCGTCT 1458
Qy 441 GlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyPheThrHisAlaAspValThr 460
Db 1459 GCGCGCCCGCTCATCAGCGCTCTCGAGACCTTCGGCGGTTCACCCACGCGCGGTACT 1518
Qy 461 AlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyr 480
Db 1519 GCGGTGCCATCGACGATCTTCAGCTCACCGGTGTTTCAACGCGGACCCAGCGTCGACTAT 1578
```

```
QY 481 AlaGluLeuAsnProSerIlelleValArgAlaGlySerPheAspProSerSerGlnPro 500
Db 1579 GCGGAATTAATCCGTCGATCATCGTTCGCGTGAAGTTTCGATCCATCCAGCGCAACCG 1638
QY 501 AsnAspArgHisValAlaPheSerThrAspGlyGlyLeuAsnTrpPheGlnGlySerGlu 520
Db 1639 AACGACAGCAGCTCGCGTTCTCGACAGCGCGCAAGAACTGGTTCCAGGCGAGCGAA 1698
QY 521 ProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheVal 540
Db 1699 CCTGCGGGGTGACAGCGCGGCACCGTCGCGCATCGCGCAGCGGTCTCGTTTCGTC 1758
QY 541 TrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrp 560
Db 1759 TGGGCTCCCGCGCATCCCGGTTCGAGCTGTGTGTACGCGAGTCGGATTTGGCAACTCCTCG 1818
QY 561 AlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLys 580
Db 1819 GCTGCTTCGCAAGGTGTTCGCCCAATGCCAGATCCCGTCAGACCGGGTGAATCCAAAG 1878
QY 581 ThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPhe 600
Db 1879 ACTTCTATGCCCTATCAATGGAACTTCTATCGAAGCAGCGGCGGTGACATTC 1938
QY 601 GlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaVal 620
Db 1939 CAACCGGTGCGCGCGGTCTTCCGAGCAGCGGTGCGGTGATGTCATGTTCCACGCGGTG 1998
QY 621 ProGlyLysGluGlyAspLeuTrpLeuAlaSerSerGlyLeuTyrHisSerThrAsn 640
Db 1999 CCTGGAAGAAGAGCGATCTGTGGCTCGCTGCATCGAGCGGGCTTTACCACTCAACCAAT 2058
QY 641 GlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGly 660
Db 2059 GCGCGCAGCAGTTGGTCTGCAATCACCGCGGTATCTCCGCGGTGAAGTGGGATTTGGT 2118
QY 661 LysSerAlaProGlySerSerTrpProAlaValPheValValGlyThrIleGlyVal 680
Db 2119 AAGTCTGCGCGCGGTGTCATACCCAGCGCTCTTTGTCGCGGACGATCGGAGCGGT 2178
QY 681 ThrGlyAlaTyrArgSerAspAspCysGlyThrThrTrpValLeuIleAsnAspGln 700
Db 2179 ACGGGGGGTACCGCTCCGACACTGTGGGACGACCTGGGTACTGATCAATGATACACAG 2238
QY 701 HisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgVal 720
Db 2239 CACCAATACGGAATTTGGGACAGCAATCACCGGTGACCAACCGAATTTACGGCGGGTG 2298
QY 721 TyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyGlyAlaProSerGly 740
Db 2299 TACATAGGCACGACGCGCGTGAATTTGATACGGGGACATTTGGTGGTCCGCGTCCGGA 2358
```

## RESULT 3

```
US-10-156-761-1845
; Sequence 1845, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
```

```
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1845
; LENGTH: 2646
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2646)
US-10-156-761-1845
```

## Alignment Scores:

```
Pred. No.: 2,58e-229 Length: 2646
Score: 2429.50 Matches: 438
Percent Similarity: 73.99% Conservative: 111
Best Local Similarity: 59.03% Mismatches: 174
Query Match: 60.20% Indels: 19
DB: 15 Gaps: 8
```

US-09-917-376-3 (1-740) x US-10-156-761-1845 (1-2646)

```
QY 6 TyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn 25
Db 115 TACAGCTGGAAGAACGCCCGCGTCGACGGCGGGTTCGTCGCCGCGCATGCTTCAAC 174
QY 26 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyMetTyrArgTrpAsp 45
Db 175 CGCTCCGAGAGAAACCTCGCTACGCCCGCACCGACATCGCGCGCGCTACCGCTGGGCC 234
QY 46 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnTrpGlyTyr 65
Db 235 GAGTCTCTCAAGACCTCGGACCGCGCTCGACTCGGCTGGAGCGACTGGGGGCGAC 294
QY 66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 85
Db 295 ACGGGTGTCTGAGCTCGCTCCGACTCGCTCGACCGCAAGGTGTACGGCGCGGTC 354
QY 86 GlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGln 105
Db 355 GGCAGCTACACGAACAGCTGGGACCGCGCAACGGTCCGCTCAGTCCGGGCGCGG 414
QY 106 GlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArg 125
Db 415 GCGCGGAGTGGCAGAAAGACCGACCTGCCCTTCAAGTGGCGGGGAACATGCCGGCGG 474
QY 126 GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAla 145
Db 475 GGCATGGGCGAGCGGCTCGCGGTGCGACCGCAAGCAAGCGTGTGTATCTCGGCGG 534
QY 146 ProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThr 165
Db 535 CCCAGCGGCAAGGGGTGTGGCGGTGCGACCGACTCGGGGGCTCTCTGGTCCAGGTCAAC 594
QY 166 AsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSer 185
Db 595 GACTTCCCGAACCTCGGCACCTACGTGAGAGACCGCACCGACAGCGGGTACCGCGTCC 654
QY 186 AspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerSerLeuGlyGlnAla 205
Db 655 GACAACAGGCGATCGTGTGGGTCACTTCGACGAGTCCGACGGGTGTCGCCGGGAGGTCC 714
QY 206 SerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSerArgAsp 225
Db 715 ACGCGGACGGTGTACGTGCGGGTTCGCCGCAAGGACAACTCCCGTCTATCGCTCCACGGAC 774
QY 226 GlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIleProHisLysGly 245
Db 775 GCGGGCGCGACCTGTGTCGGGTGGCGCGCCACCGGCCCATCTCGGCCACAGAGGC 834
QY 246 ValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyr 265
Db 835 GTGCTGGAGCGCGGAACGGCTGTGTACCTCGCTCGGTACAGCAAGGGCGGACCGGTAC 894
QY 266 AspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrpThrArgIle 285
```



Db	2277568	CGCTCCGAGAGAACTCGCTACGCCCGACCGACATCGCGCGCTACCGCTGGGC	2277627
QY	46	AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr	65
Db	2277628	GAGTCTCTCAAGACCTGGACCGCGCTGCTCGACTCGGTGGAGCGACTGGGGGCAC	2277687
QY	66	AsnGlyValValSerIleAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal	85
Db	2277688	ACGGGTGTCTGAGCGCTCGCCCTCCGACTCGGTGACCCCGAAACAGAGTGTACGCGCGCGTC	2277747
QY	86	GlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGln	105
Db	2277748	GGCAGCTACAGCAAGCTGGGACCCGGGACGAGGTGCGTCTAGGTCCGGCGACCGG	2277807
QY	106	GlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArg	125
Db	2277808	GGCGCGAGCTGGCAGAAAGACCGACCTGCGCTTCAAGCTGGCGGGAAACATGCGCGCGCG	2277867
QY	126	GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAla	145
Db	2277868	GGCATGGCGGAGCGGTCTCGCGGTGCGACCCGAAACAGGAGTGTGTATCTCGCGCGG	2277927
QY	146	ProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyValaThrTrpSerGlnMetThr	165
Db	2277928	CCAGCGGCAAGGGGTGTGGCGGTGACGAGCTCGGGGCGCTCTGTGTGCAAGGTAC	2277987
QY	166	AsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSer	185
Db	2277988	GACTTCCCAGACGTGGCACCTACGTGAGGAGCGGACCGACGACGAGGGGTACGCGTCC	2278047
QY	186	AspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerLeuGlyGlnAla	205
Db	2278048	GACAAACCGAGGATCGTGTGGGTGACCTTTCAGAGTTCGACGGGTGCGGGGAGCTCC	2278107
QY	206	SerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSerArgAsp	225
Db	2278108	ACCGGACGGTGTAGTCTGGGTGCGCGGACAGGACACTCCGTCTATCGTCCACGGAC	2278167
QY	226	GlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIleProHisLysGly	245
Db	2278168	CGGGCGCGACCTGGTCCCGGTGGCGCGCCAGCCCGCCATCTCCGCCACAGGGC	2278227
QY	246	ValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyr	265
Db	2278228	GTGCTGGACCGCGGCAACGGCTGTCTGTACTCTCGGTACAGCGCAAGGGCGGACCGTAC	2278287
QY	266	AspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrpArgIle	285
Db	2278288	GACGGCGGCAAGGACAGCTGTGGCGGTACACGAGACCGCGGACCTGGAGCAATC	2278347
QY	286	SerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIle	305
Db	2278348	AGCCCGGTTCGGAGGCGGACACC-----TACTACGGCTTCAGCGGGTACCGGTG	2278398
QY	306	AspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrpProAspThr	325
Db	2278399	GACCGGACATCCGGGACCGGTGTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	2278458
QY	326	IleIlePheArgSerThrAspGlyAlaThrTrpThrArgIleTrpAspTrpThrSer	345
Db	2278459	CAGCTCTTCGCTCCAGGACAGCGGCGACCTCGGAGGAGGCTGGGACTACACCTCG	2278518
QY	346	TyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpLeuThrPhe	365
Db	2278519	TATCCGAGCGCTCGAACCGCTTCACCATGGATGTCTGTCTCGCCCTGGCTACCTGG	2278578
QY	366	GlyValGlnProAsnProProValProSerProLysLeuGlyTrpMetAspGluAlaMet	385
Db	2278579	GGAGCGAACCCCGCACCGCGGACAGACCCCGAACTCGGTGTGATGACCGAGTCCCTG	2278638
QY	386	AlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAla	405
Db	2278639	GAGATCGACCCCTTCGACTCCCGCGGCGATGATGATGATGATGATGATGATGATGATG	2278698
QY	406	ThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValLys	425
Db	2278699	ACGGACAACTGACGAACCTGGGACAGCGGAAGCAGTTCACCATCAAGCCGATCGCGCG	2278758
QY	426	GlyLeuGluGluThrAlaValAsnAspLeuIleSerProProSer-----GlyAlaProLeu	444
Db	2278759	GGCTCGAGAGAGAGCGCGCTCAACGACCTCGCTCGCTCCCTCGCGCGCGCCAGCTG	2278818
QY	445	IleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaValProSer	464
Db	2278819	TTCAGCGCGCTCGGTGACATCGCGCGGTTCGGGACACGAGACCTCACACAGGTGCCGTG	2278878
QY	465	ThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeuAsn	484
Db	2278879	CTGATGTACAGCTCGCGGAATTCACACGAGCACCAGCTCGACTACGCGGACCGAC	2278938
QY	485	ProSerIleIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArgHis	504
Db	2278939	CGGGGACCGGTGTGGGTGCGGCAATCTCGAC-----TCGGGTCCG-----CAT	2278983
QY	505	ValAlaPheSerThrAspGlyLysAsnTrpPheGlnGlySerGluProGlyGlyVal	524
Db	2278984	GTGGCGTCTTCGACGCAACACGCGGCACTGGTTCCGGGGGGGGGAGCCCTTCGGGGGT	2279043
QY	525	ThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaProGly	544
Db	2279044	AGCGGGGTGGGACGCTCGCGCGCGTCCGACGCGCAGTCTGTTCTGTTGAGCCCGCG	2279103
QY	545	AspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGln	564
Db	2279104	GGCACCGGG-----GTGCAGTACACGACCGGGTTCGGCACCTCGTGTGCGGTCCGCG	2279157
QY	565	GlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrAla	584
Db	2279158	GGCTCTCCCGCGCGGGCGATGCTGAGTCCGACCGGTTCGCGCGTCCGCGGAGACCTT	2279217
QY	585	LeuSerAsnGlyThrPheTyrArgSerThrAspGlyValThrPheGlnProValAla	604
Db	2279218	TTCAAGTCCGGCAGGTCTACGTCACTCGGACGCGCGGGCGACCTTCACGGGTCCGCG	2279277
QY	605	Ala-----GlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaValProGlyLys	623
Db	2279278	GGCACCGGCTGCGGAGCGCGGACAGC-----GTGCGCTTCAAGGCGCTGCCCGGACG	2279331
QY	624	GluGlyAspLeuTrpLeuAlaAlaSerSer-----GlyLeuTyrHisSer	638
Db	2279332	AAGGGCGACATCTGGCTGGCGCGCGCGGACGCGCGGTACGGGCTGTGGCACTCG	2279391
QY	639	ThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGly	658
Db	2279392	ACGGACGCGCGCGCGCTTCACCAAGCTCGCCACCGTCCGACCGCGCGACCATCGCG	2279451
QY	659	PheGlyLysSerAlaProGlySerSerTyrProAlaValPheValValGlyThrIleGly	678
Db	2279452	TTTCGCAAGCGCGGACCGCGCTCTACACAGCTCTTACACAGCGCGGAGATCGCG	2279511
QY	679	GlyValThrGlyAlaTyrArgSerAspCysGlyThrThrTrpValLeuIleAsnAsp	698
Db	2279512	GGTGTGCGGGCATCTTCGGTTCGACCGCAAGGGCGGAGCTGGACCCCGCTCAACGAC	2279571
QY	699	AspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArg	718
Db	2279572	GATGCCACCATGGGGTGGACGGCGCGCGATCACCGGTGACCCCGAGGTCTACGGG	2279631
QY	719	ArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyAlaPro	738
Db	2279632	CGCGTGTGTGTGCGAGCAAGCGGCGCGGATCGTCTACGGCGACACCGCGGCTCTCG	2279691
QY	739	SerGly	740
Db	2279692	GACGGC	2279697

## RESULT 5

US-10-026-994-4  
 ; Sequence 4, Application US/10026994  
 ; Publication NO. US20030113732A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dunn-Coleman, Nigel  
 ; APPLICANT: Goedegebuur, Frits  
 ; APPLICANT: Ward, Michael  
 ; APPLICANT: Yao, Jian  
 ; TITLE OF INVENTION: EGVI Endoglucanase and Nucleic Acids  
 ; FILE OF INVENTION: Encoding the Same  
 ; FILE REFERENCE: GC698  
 ; CURRENT APPLICATION NUMBER: US/10/026,994  
 ; CURRENT FILING DATE: 2002-04-30  
 ; NUMBER OF SEQ ID NOS: 4  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 4  
 ; LENGTH: 2517  
 ; TYPE: DNA  
 ; ORGANISM: Trichoderma reesei  
 US-10-026-994-4

Alignment Scores:  
 Pred. No.: 1,78e-145 Length: 2517  
 Score: 1579.00 Matches: 325  
 Percent Similarity: 59.26% Conservativity: 123  
 Best Local Similarity: 42.99% Mismatches: 264  
 Query Match: 39.12% Indels: 44  
 DB: 15 Gaps: 20

US-09-917-376-3 (1-740) x US-10-026-994-4 (1-2517)

QY 6 TyrTrpSerAsnValAlaIle---GlyGlyGlyGlyPheValAspGlyIleValPhe 24  
 DB 61 TTTTCAGGAACAGCTCAAGCTCGCGCGCGCGCGCTTCGCCCGCATCATCTTC 120  
 QY 25 AsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrp 44  
 DB 121 CATCCCAAGACAAAAGGCGTAGCATATGCACGAACAGATATTGGCGGCTGTACCGCTC 180  
 QY 45 AspAlaIleAsnGlyArgTrpIleProLeuLeuAspTrpVal-----GlyTrp 60  
 DB 181 AAC--GCCGACGACTCATGACCGCGCTCAGGATGGGATTGTGTAATGCCGCTGG 237  
 QY 61 AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAspProIleAsnThrAsnLys 80  
 DB 238 CACAACCTGG-----GGCATCAGCTGTGCGCTTGATCCGAGCATCAAAAG 288  
 QY 81 ValTrpAlaIleValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 100  
 DB 289 GGTATGCCGAGTCGGCATGTATACGAACAGCTGGGATCCGAGTAATGGAGCCATCAT 348  
 QY 101 ArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGly 120  
 DB 349 CGCTCGTCAGACCGCGCGCAACGTGGTCTTCAACCACTTGGCTTCAAAAGTCGGGGT 408  
 QY 121 AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle 140  
 DB 409 AACATGCCAGACCGGAGCGGAGAGCTGTGGCTGTGCTATCCGCCCAATCCCAACATC 468  
 QY 141 LeuTyrPheGlyAlaProSerGlyLysGlyLeuTyrArgSerThrAspSerGlyAlaThr 160  
 DB 469 ATCTACTTTGGTGTCTCGTCTAGGAACCGGCTCTGGAAAGTCTACGGACGGCGGTGACC 528  
 QY 161 TrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThr 180  
 DB 529 TTTTCAAGGTCCTCGTCTCAGGAACCGGCTCTGGAAAGTCTACGGACGGCGGTGATTCC 588  
 QY 181 ThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSer 200  
 DB 589 AACGGCTACAAACAGCGAACAAGGACTCATATGGTGGGTACGTTTCGATCAACACGACG 648  
 QY 201 SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspPro---AsnAsnPro 219

DB 649 ACACCGGGGAGCCACGCTCTCGTATCTTTGTTGGACCGCTGATAACTACTGCTTCA 708  
 QY 220 ValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGly 239  
 DB 709 GTCTATGTGACACGAATGCGGCTCCAGCTGGAGTGTGTACCGGGGAGCCAGGAAA 768  
 QY 240 PheIleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSer 259  
 DB 769 TACTTTCCTCACAAGCGGAAACTGCGAGCCAGCAGAGAGCGCTTGTATCTGACTTATCC 828  
 QY 260 AsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSer 279  
 DB 829 GATGGCACAGGCGCGTATGATGGCACACTTGGCTCAGTGTGGAGTACGACATTGCGAGG 888  
 QY 280 GlyThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGly 299  
 DB 889 GGAACCTTGGAAAGACATCACCCCTGTCTGTGATCAGATCTA-----TACTTTGGC 939  
 QY 300 TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIle 319  
 DB 940 TTTGGCGGCTTGGCTCGATTGCAAAAGCCAGGAACCCCTTGTCTTCTTTGAAC 999  
 QY 320 SerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArg 339  
 DB 1000 TCTTGGTGGCCAGATGCTCAGCTGTTTGGTGGCCGACTCTGGGACAAATGGAGCCG 1059  
 QY 340 IleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAla 359  
 DB 1060 ATCTGGGCGTGGCGAGCTATCCGACTGAGACCTATTACTAGCAGATCTCAACTCCCAAA 1119  
 QY 360 GluProTrpLeuThrPheGly---ValGlnProAsnProProValProSer----- 375  
 DB 1120 GCACCGTGTATCAAGAACAACTTTATCGATGTGACGCGGAGTCACTCGGTCGTCTC 1179  
 QY 376 ---ProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArg 394  
 DB 1180 ATCAAGCGCTCGGCTGGATGATTGAGTCTCTCGAGATTGACCAACCCAGCAGCAACCAC 1239  
 QY 395 MetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSer 414  
 DB 1240 TGGCTCTACGCGACCGGAATGACATCTTTGGCGGCCAGTCTCACCACATGGGACAG 1299  
 QY 415 GlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluGluThrAlaValAsnAsp 434  
 DB 1300 GCCACATGTCTCAATCAATCACTGGCAGACGGCATCGAGGAATCTCCGTCAGAGAC 1359  
 QY 435 LeuIleSerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPhe 454  
 DB 1360 CTGGCTCTGCACCGCGGGAAGCGAGCTATTGGCGCGAGTCCGAGACGACACACGCTTC 1419  
 QY 455 ThrHisAla-----AspValThrAlaValProSerThrIlePheThrSerProVal 471  
 DB 1420 ACCTTTGGCAGCAAAACGACCTCGGAGCATCGCGCAGACGGCTCTGGGACGCGCCACA 1479  
 QY 472 PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAla 491  
 DB 1480 TGGGCCACTCGACGAGCTCGACTACCGCGGAACCTCGTCAAGAGCGTCTGTCGCCGTC 1539  
 QY 492 GlySerPheAspProProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly 511  
 DB 1540 GGCAACACCGCGGACCAACAG-----GTGGCCATCTCGTCCGACGGC 1584  
 QY 512 GlyLysAsnTrpPheGlnGlySerGluProGlyGlyValThrThrGlyGlyThrValAla 531  
 DB 1585 GCGCGCAGTGGAGCATCGACTACCGCGCGCACAGCTCCATGAACCGCGGACGCGTGGCC 1644  
 QY 532 AlaSerAlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValVal 551  
 DB 1645 TATTGGCGCGACGGGACACGATCTCTGTGTCAGCGCCTCGTCCGCG-----GTG 1695  
 QY 552 TyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGln 571



```
Db 1696 CAGCGCTCGAGTTCACGGGAGCTTTGCTCGCTCGAGCTCCCGCGGGCGCCGTC 1755
Qy 572 IleArgSerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyr 591
Db 1756 ATCGCTCGGACAGAGAACAGCAACAGCGCTTCTACGCGGCTCGGATCGACCTTTTAC 1815
Qy 592 ArgSerThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGly 611
Db 1816 GTCAGCAAGAGCACCGGAGCAGCTTC-----ACGCGGGG---CCCAAGCTGGGC 1863
Qy 612 AlaValGlyValMet-----PheHisAlaValProGlyLysGluGlyAspLeuTrpLeu 629
Db 1864 AGCGAGGAGCATCCGGATATCCTCTCACCAGCACCGCGGGCAGCTTGATGTC 1923
Qy 630 AlaLaserSerGlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIle--- 648
Db 1924 TCGACCGAGCTCGCATATTCCGCTCCACAGACTCGGCGACGACCTTTGGCCAGTCTCC 1983
Qy 649 ThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyr 668
Db 1984 ACCGCTCGACCAACACCTACCAGATCGCCCTGGGTGGGTCTCA--GGCTCGAACTGG 2040
Qy 669 ProAlaValPheValGlyThrIleGlyGlyValThrGlyAla-----TyrArgSer 686
Db 2041 ---AACCTGTATGCTTCGGCACC-----GCCCGCTCAGGCGCTCGCCTCTACGCCAGT 2091
Qy 687 AspAspCysGlyThrThrTrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrp 706
Db 2092 GGAGACAGCGCGCTCTCGGAGGACATCCAGGGCTCCAGGGCTTCGGCTCCATCCAGC 2151
Qy 707 GlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgArgValTyrIleGlyThrAsnGly 726
Db 2152 AGCACCAAGGTCCGCGGAGCGGAGCACCGCGCGGCAAGTCTACGTGGGCAACCAACGCG 2211
Qy 727 ArgGlyIleValTyr-----GlyAspIleGlyGlyAlaProSerGly 740
Db 2212 CGGGCGCTCTTTTACGCTCAGGGAACCTCGCGCGCGCGACGCGGGCGG 2259

RESULT 6
US-10-026-994-1
; Sequence 1, Application US/10026994
; Publication No. US20030113732A1
; GENERAL INFORMATION:
; APPLICANT: Dunn-Coleman, Nigel
; APPLICANT: Goedegebuur, Frits
; APPLICANT: Ward, Michael
; APPLICANT: Yao, Jian
; TITLE OF INVENTION: EGGV Endoglucanase and Nucleic Acids
; FILE REFERENCE: GC698
; CURRENT APPLICATION NUMBER: US/10/026,994
; CURRENT FILING DATE: 2002-04-30
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2710
; TYPE: DNA
; ORGANISM: Trichoderma reesei
US-10-026-994-1

Alignment Scores:
Pred. No.: 1,97e-145 Length: 2710
Score: 1579.00 Matches: 325
Percent Similarity: 59.26% Conservative: 123
Best Local Similarity: 42.99% Mismatches: 264
Query Match: 39.12% Indels: 44
DB: 15 Gaps: 20

US-09-917-376-3 (1-740) x US-10-026-994-1 (1-2710)

Qy 6 TyrThrTrpSerAsnValAlaIle---GlyGlyGlyPheValAspGlyIleValPhe 24
Db 101 TTTTCATGGAAGACGTCAAGCTCGCGCGCGCGGCTTCGTCGCCGGCATCATCTTC 160
```



Db 541 ---CGGCACGCGGCTGCTCAAGTCGACCGGCGGCGCCACTTGGCGACCGCGGACC 597  
QY 166 AenPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSer 185  
Db 598 GCTTTCGGG-----GCGAAGCGGAACCTCC 621  
QY 186 AspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerLeuGlyGlnAla 205  
Db 622 TCCGGGACGGAGTCTGTGTC-----CTCGTCGCCGCC 654  
QY 206 SerLysThrIlePheValGlyValAlaAspProAsnAsnPro-----Val 220  
Db 655 GGGCGACCGTCTACCGCGCTGGGTGACCGCGGACCGCACTCGGGCAGCGGACCTG 714  
QY 221 PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe 240  
Db 715 TACCGCACGCGCGAC---GGCACGACCTGGGGCGCTCCCGCGCGGCGCTCCGCGCAC 771  
QY 241 -----IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAla 257  
Db 772 TCCGCCAAGTCCCGCTCCGCGCGCTGACGACGACGACGACCGCGGAGCTGTACGTGACG 831  
QY 258 ThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerVal 277  
Db 832 TACGGCGACGACCGCGCGCGCGCGCGCTCGACGCGAGGTGCACAACTGCGTAC 891  
QY 278 ThrSerGlyThrTrpThrArgIleSerProVal-----ProSerThrAspThr 293  
Db 892 GCCCGGGAGCTGACCGAGGTACCCCGGTGAAGCGCGGGGACGACGACGACGCGC 951  
QY 294 AlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIle 313  
Db 952 TCGGGCGACACCTTCGCTTACGCGCGGGTTCGCGTGCAGCGCGCGCGCGCGCACCTC 1011  
QY 314 MetValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAspGly 333  
Db 1012 GTGCTCTCACCAACACCGCTGGCGCGGACGCGACCGGTCTTCGGTCCACGACGCGC 1071  
QY 334 GlyAlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyr 353  
Db 1072 GGCCGTACTGACGCTCCTCAAGGAC-----GCCGCG 1104  
QY 354 ValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProVal 373  
Db 1105 GTGTTTCGAGTCTCGAGACTCCCTCTCTCGACTGGGCGGACGAC----- 1149  
QY 374 ProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAsp 393  
Db 1150 ---AAGCCGAAGTTCGGCTGGTGGATCCAGGCGCTCGCGGTTCGACCCGTACACCTCC 1206  
QY 394 ArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAsp 413  
Db 1207 CACGTCTGTACGGACCGCGCGGACCTCTACGGCACCGGACCTCAAGCGCTGG--- 1263  
QY 414 SerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluThrAlaValAsn 433  
Db 1264 -----GCACCGCGGATCCGCGGCTGGAGAGAGCGCGCTGGCG 1302  
QY 434 AspLeuIleSerProProSerGly---AlaProLeuIleSerAlaLeuGlyAspLeuGly 452  
Db 1303 CAACCTGATCTCGCCCCCGTTCGGGGAGGACACCTGATCAGCGGACTCGGGGACATCG 1362  
QY 453 GlyPheThrHisAlaAspValThrAlaValProSerThr---IlePheThrSerProVal 471  
Db 1363 GTGATGTACACGACGCGCTCAGCGGTCTCGTCCGCGGATCGGCGACGACCCCGG 1422  
QY 472 PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAla 491  
Db 1423 TTCGGTTCGGGACGGGACTCGCGGAGCGCGGCGGCTGATGCTGCTCGCACG 1482  
QY 492 GlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly 511  
Db 1483 GGCTGGGCGCACCGGCAACGCG-----GCGTACTCCACGACGCGC 1524

QY 512 GlyLysAsnTrp-----PheGlnGlySerGluProGlyGlyValThrThrGlyGlyThr 529  
Db 1525 GGGCGACCTGGCGCGCTTCGAGCGCCAGCCGACATCCCAAGGACGACCGACCGGACCG 1584  
QY 530 ValAlaAlaSerAlaAspGlySerArgPheValTrpAla-----ProGlyAsp 545  
Db 1585 ATCGCCACAGTTCGCGACGCGCGCACATGCTGTGTCTTCTGTGCACTGGAGCGCG--- 1641  
QY 546 ProGlyGlnProValValTyrAla-----ValGlyPheGlyAsnSerTrpAla 561  
Db 1642 -----ACGACGTACGCGGCGCCACCGCTCGACGAGCAACGCGCGGAGTGTGTC 1689  
QY 562 AlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThr 581  
Db 1690 GAGGTCTCTCTTCCGAAAGGCGCGCACACCGGTGCGCGCGCGGCGGATCCGACGCGC 1749  
QY 582 PheTyrAla-----LeuSerAsnGlyThrPheTyrArgSerThrAspGlyValThr 599  
Db 1750 TTCTACGCTACGACTTCGACCAATGGAACGCTATACGCCAGCACTGACAGTGGCGCTCG 1809  
QY 600 PheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAla 619  
Db 1810 TTCACGCGCGTGGCGGCGGACTGCCC---TCCGCGACAGCCAGTTCAGCTGGTGGCG 1866  
QY 620 ValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSer---SerGlyLeuTyrHisSer 638  
Db 1867 GCGCGGGACGAAGCGCGGACCTGTGCTCTCCGCCAAATGGAACGCGGCTCTACCGTCC 1926  
QY 639 ThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGly 658  
Db 1927 ACCGACGCGGGGACACCTTCGCGAGGATCGACAGCTGTGGGCGCTCGTACACCTCGGC 1986  
QY 659 PheGlyLysSerAlaProGlySerSerTyrProAlaValPheValValGlyThrIleGly 678  
Db 1987 TTCGGCAAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2046  
QY 679 GlyValThrGlyAlaTyrArgSerAspAspCysGlyThrThrTrpValLeuIleAsnAsp 698  
Db 2047 ACCATCACCCTGCTTACCGCTCCGACGCGCGCGGCGGCGGCGGCGGCGGCGGCGG 2106  
QY 699 AspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArg 718  
Db 2107 GACGCGCACAGTGGGGGTGGATCGCGGAGGCGCGTCTCGTGACCGCGCATCCACGCG 2166  
QY 719 ArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAsp 733  
Db 2167 CGGGTCTACCTCGCCACCAACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2211

## RESULT 8

US-10-156-761-1/c  
; Sequence 1, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156, 761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 1  
; LENGTH: 9025608  
; TYPE: DNA

; ORGANISM: Streptomyces avermitilis  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (4187715)  
 ; OTHER INFORMATION: a, t, c, g, other or unknown  
 US-10-156-761-1

## Alignment Scores:

Pred. No.: 1,24e-123 Length: 9025608  
 Score: 1405.00 Matches: 308  
 Percent Similarity: 54.57% Conservative: 104  
 Best Local Similarity: 40.79% Mismatches: 257  
 Query Match: 34.81% Indels: 86  
 DB: 15 Gaps: 21

US-09-917-376-3 (1-740) x US-10-156-761-1 (1-9025608)

Qy 6 TyrThrTrpSerAsnValAlaIleGlyGlyPheValAspGlyIleValPheAsn 25  
 Db 3161882 TACGGCTGGCGCAACCGCGTCATCGGGGGCACCGGCTTCGTACCGCGGTCTCTCCAC 3161823  
 Qy 26 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrpAsp 45  
 Db 3161822 CCCTCGGTACGGGTCTCGCTACGCCCGGACCGACATCGCGCGCGCTACCGCTGGAC 3161763  
 Qy 46 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr 65  
 Db 3161762 GACCGCGCGCGCGTGGACCGCGCTCATCGACCACTCGGCTGGGACGACTGGAACTC 3161703  
 Qy 66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 85  
 Db 3161702 CTCGGCGTCGAGCGATGGCGGTGCGACCCACGACCGGACCGGCTCTACCTCGCGGTG 3161643  
 Qy 86 GlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGln 105  
 Db 3161642 GGCACCTACGCCCGATGCGTGGCGGGCAAC---GGCGGTCCTCGCGTCCGAGACCGC 3161586  
 Qy 106 GlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArg 125  
 Db 3161585 GCGGCCACTGGACCGCGACCGACTGACCGTGAAGCTCGCGGCAACGAGGACGGCGC 3161526  
 Qy 126 GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAla 145  
 Db 3161525 GCGCGCGGTGAGCGACTCTCTGTCGACCGCGCGACGACACCTCTGCGTGGCGACG 3161466  
 Qy 146 ProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThr 165  
 Db 3161465 ---CGGCACGACGGGTCTCAAGTCGACCGACCGCGGCGCGCTCGGCGACCGCGAC 3161409  
 Qy 166 AsnProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSer 185  
 Db 3161408 GCCTTCCCG-----GGGAGCGGAATCC 3161385  
 Qy 186 AspileGlnGlyValValTrpValAlaPheAspLysSerSerSerLeuGlyGlnAla 205  
 Db 3161384 TCCGGCAGGAGTCGTTC-----CTCGTCCGCC 3161352  
 Qy 206 SerLysThrIlePheValGlyValAlaAspProAsnAsnPro-----Val 220  
 Db 3161351 GGGCGCACCGTCTACCGCGGTGGGTGACGCGACGCGCTCGGGCGCGCAACCTG 3161292  
 Qy 221 PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe 240  
 Db 3161291 TACCGCAGCGCGAC---GGCACACCTGGGGGGCGGTCCCGCGCGCGCTCGGCGAC 3161235  
 Qy 241 -----IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAla 257  
 Db 3161234 TCCGCCAAGTCTCCGCTCCGCGCGGTGACACGCGACACCGCGGAGCTGTAGTGACG 3161175  
 Qy 258 ThrSerAsnThrGlyProTyrAspGlySerGlyAspValTrpLysPheSerVal 277  
 Db 3161174 TACGCGACGACCGCGCGCGCGCGCGTCCGACGCGACGCGTGCACAAAGCTGGTACC 3161115

Qy 278 ThrSerGlyThrTrpThrArgIleSerProVal-----ProSerThrAspThr 293  
 Db 3161114 GCCACGGGACGTGACCGAGGTCAACCGGTGAGCGCGGCGGACGACGACGACGCGC 3161055  
 Qy 294 AlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIle 313  
 Db 3161054 TCGCGGACACCTTCGCTACGGCGGGTTCGCGCTGACGCGCGCGCGCGCCGCCCTC 3160995  
 Qy 314 MetValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAspGly 333  
 Db 3160994 GTCGTCTCCACCAACACCGCTGGCGCGACGCGACGCGTCTTCGCTCCACGACGCGC 3160935  
 Qy 334 GlyAlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyr 353  
 Db 3160934 GCGGTATCTGGACGTCCTCAAGGAC-----GCCGCC 3160902  
 Qy 354 ValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProVal 373  
 Db 3160901 GTGTTCGACGTGTCGAGACTCCCTTCCTCGACTGGGCGACGAC----- 3160857  
 Qy 374 ProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAsp 393  
 Db 3160856 ---AAGCGAAGTTCGGTGGTATCCAGCGCTCGCGGTGACCCGTCACGATCCCG 3160800  
 Qy 394 ArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAsp 413  
 Db 3160799 CACGTGTTGTCGCGACCGCGCGACCTCTACGCGACCCCGGACCTCAAGCGCTCG--- 3160743  
 Qy 414 SerGlyGlnIleHisIleAlaProMetValLysGlyLeuGluGluThrAlaValAsn 433  
 Db 3160742 -----GCACCGGATTCGCGCGCTGGAGAGAGCGCGTGGCGC 3160704  
 Qy 434 AspileIleSerProSerGly---AlaProLeuIleSerAlaLeuGlyAspLeuGly 452  
 Db 3160703 CAACTGATCTCGCCCCCGTGGGAGGACACCTGATCAGCGGACTCGGAGCATCGGT 3160644  
 Qy 453 GlyPheThrHisAlaAspValThrAlaValProSerThr---IlePheThrSerProVal 471  
 Db 3160643 GTGATGTACACGACGCGCTCACGCGCTCTCGTTCGCGGCGGATGCGCGAACCCCGTG 3160584  
 Qy 472 PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAla 491  
 Db 3160583 TTCGGTTCGCGACGCGGACTTCGCGAGCGCGCGCGCGCGGTATGTCGTCCGACG 3160524  
 Qy 492 GlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly 511  
 Db 3160523 GGCTGGGCGACCCCGCAACGCGC-----GCGTACTCCACGACGCGC 3160482  
 Qy 512 GlyLysAsnTrp-----PheGlnGlySerGluProGlyGlyValThrThrGlyGlyThr 529  
 Db 3160481 GGGCGGACCTGGCGCCCTTCGAGGCCCGACGCGCATCGCCCAAGGACGACCGGGACCG 3160422  
 Qy 530 ValAlaAlaSerAlaAspGlySerArgPheValTrpAla-----ProGlyAsp 545  
 Db 3160421 ATCGCACAGTTCGCGACGCGCGCACACTGCTGTGTCTCTCGTGCACCTGGGACGCGC--- 3160365  
 Qy 546 ProGlyGlnProValValTyrAla-----ValGlyPheGlyAsnSerTrpAla 561  
 Db 3160364 -----ACGACGTACCGCGCCCGCCCGTTCGACGAGCAACCGCGCGAGCTGTCC 3160317  
 Qy 562 AlaSerGlnGlyValProAlaAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThr 581  
 Db 3160316 GAGGTCTCTCTCCGAGGGCGCGCACCGGTTCGCGGACCGCGCGCGATCCGACGCGC 3160257  
 Qy 582 PheTyrAla-----LeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThr 599  
 Db 3160256 TTCTACGCTACGACTTCGACAAATGGAACGCTATACCGCGACGACTGACAGTGGCGCTCG 3160197  
 Qy 600 PheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAla 619  
 Db 3160196 TTCACGGCGCGTGGCGGCGACTGCC-----TCGCGACGACGCGAGTTCAGGTGTCGCG 3160140  
 Qy 620 ValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSer---SerGlyLeuTyrHisSer 638

Db 3160139 GCGCCGGGACGAGCGGCGCTGTGGCTCTCCGCCAAATGAAACGGGCTCTACCGGTCC 3160080  
QY 639 ThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGly 658  
Db 3160079 ACCGAGCGGGGACACACTTCGCCAGGATCGACAGCTGCTGGGCTCGTACACCTCGGC 3160020  
QY 659 PheGlyLysSerAlaProGlySerSerTrpProAlaValPheValGlyThrIleGly 678  
Db 3160019 TTCCGCAAGCGCGCGACGCGCGCGACTACCGCGCGATCTACCGAGTCTCGCTCGAGGAG 3159960  
QY 679 GlyValThrGlyAlaTyArgSerAspCysGlyThrTrpValLeuIleAsnAsp 698  
Db 3159959 ACCATCACCGCGCTTACCGCTCCGACGACGCGCGAGGATGGGTCCGGATCAACGAC 3159900  
QY 699 AspGlnHisGlnTyArgSerTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArg 718  
Db 3159899 GACGCCACCACTGGGGTGGATCGCGAGGCGCTCGGTGACCCCGCATCCAGCGC 3159840  
QY 719 ArgValTyIleGlyThrAsnGlyArgGlyIleValTyGlyAsp 733  
Db 3159839 CGGGTCTACTCCGCCAACCGCGCGCGCATCCAGTACCGGAG 3159795

## RESULT 9

US-09-927-827-21  
; Sequence 21, Application US/09927827  
; Publication NO. US2003036176A1  
; GENERAL INFORMATION:  
; APPLICANT: Bower, Stanley G.  
; APPLICANT: Ramseier, Thomas M.  
; TITLE OF INVENTION: Directed Genetic Engineering of Xanthomonas campestris  
; FILE REFERENCE: 38-10(15824)B  
; CURRENT APPLICATION NUMBER: US/09/927,827  
; PRIOR FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: US 60/279,493  
; PRIOR FILING DATE: 2001-03-28  
; NUMBER OF SEQ ID NOS: 69  
; SEQ ID NO 21  
; LENGTH: 3668  
; TYPE: DNA  
; ORGANISM: Xanthomonas campestris  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1001)..(2668)  
US-09-927-827-21

## Alignment Scores:

Pred. No.: 7,17e-102 Length: 3668  
Score: 1139.00 Matches: 267  
Percent Similarity: 50.85% Conservative: 120  
Best Local Similarity: 35.09% Mismatches: 295  
Query Match: 28.22% Indels: 79  
DB: 10 Gaps: 21

US-09-917-376-3 (1-740) x US-09-927-827-21 (1-3668)

QY 1 AlaThrTrpGlnProTyTrpTrpSerAsnValAlaIleGlyGlyGlyPheValAsp 20  
Db 1217 GCCAGTCCGGCCCTACCAAGTGGCGAGTTCGCATGCGGTGGCGGCTTTGTACCC 1276  
QY 21 GlyIleValPheAsnGluGlyAlaProGlyIleLeuTyTrpValArgThrAspIleGly 40  
Db 1277 GGTGTGCTTTTCATCCCGCCGAACTGGTCTGGCTATCGCGCACCATGATGGGTGGC 1336  
QY 41 MetTyArgTrpAspAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrp 60  
Db 1337 GGTACCGCTGGGATGCGGAGCGGACGAGTGGAGCCGGCTGACCGACTGGTGGCGCT 1396  
QY 61 AsnAsnTrpGlyTyAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 80  
Db 1397 GACGATGAACCTGATGGGATCGACGATTCGCGCTGACCCCGCCGATCCGATGCG 1456  
QY 81 ValTrpAlaAlaValGlyMetTyTrpAsnSerTrpAspProAsnAspGlyAlaIleLeu 100

Db 1457 CTGTATCTCGCGCGCGCACTATATGTCATGAA---CGCGCCGGCACTGCGCGAGTGTG 1513  
QY 101 ArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGly 120  
Db 1514 CGTCTGTTCAACCGCGCGCGCAGTTCGAGCGTCCGACCTGCGTTTAAGCTGGGTGGT 1573  
QY 121 AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle 140  
Db 1574 AACCACTGGCGCGCGCAATGCGAGCGGCTGGCGTGACCGCACAGTATGGCGCGGTG 1633  
QY 141 LeuTyPheGlyAlaProSerGlyLysGlyLeuTyArgSerThrAspSerGlyAlaThr 160  
Db 1634 CTGCTCTGGGCTCGCGCGATGCC---GGCTGTGGCGTAGCGACGATCGCGCGCGCAC 1690  
QY 161 TrpSerGlnMetTrpAsnPheProAspValGlyThrTyIleAlaAsnProThrAspThr 180  
Db 1691 TGGCGCAAGGTGGCGTTCGCGAGCGCGCGCTGGCGGTCACCGCGCGCAATCAT 1750  
QY 181 ThrGlyTyArgSerAspIleGlnValValTrpValAlaPheAspLysSerSerSer 200  
Db 1751 GTTGGCGCGAGCAGCGCGGTG---GGATCGCTTTGTCGTTCGACGCGCAGTGGC 1807  
QY 201 SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProVal 220  
Db 1808 AACACCGGTTCGCCAACACCGCGCATCTACGTGGCGGTGCCACCGAACACGACGCTG 1867  
QY 221 PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe 240  
Db 1868 TATGTGTCGAAGATCGCGCGCGAGTGGCGACCGGTGGCGGCGGACACCGCGCGCTG 1927  
QY 241 IleProHisLysGlyValPheAspProValAsnHisValLeuTyIleAlaThrSerAsn 260  
Db 1928 CGCCCGAGCACATGCGCGCGCGCAGCATGGGCAC---TGGTATCTGAGTATGGCGAC 1984  
QY 261 ThrGlyGlyProTyArgSerGlySerGlyAspValTrpLysPheSerValThrSerGly 280  
Db 1985 CAGCCCGCGCGAGCCTGATGGCGGGGAGCCTTGTGAAATTCACGCGCGCACAGGG 2044  
QY 281 ThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyPheGlyTy 300  
Db 2045 CGTGTGCGTGTGATGACCGCGATTCGCGAG---CCAGCCAGTGGCGGTGATTCGGTGG 2101  
QY 301 SerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSer 320  
Db 2102 GGTGCGGTGGCGGTGATCGCAACATCGCAGGTGCTGCGCGCACCTTCCTCGCGGT 2161  
QY 321 TrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIle 340  
Db 2162 CGCACCGCGCGCAGCAGCTGTATCGCAGCGTGGATGGTGGCAAGCACATGACCGCGTGG 2221  
QY 341 TrpAspTrpThrSerTyProAsnArgSerLeuArgTyValLeuAspIleSerAlaGlu 360  
Db 2222 -----TTGGCGCATGCGGTGTTCGATCAGCGCGCGCG 2254  
QY 361 ProTrpLeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGlyTrp 380  
Db 2255 CCGTGG---ACCGCACATGCCCGCGCAC-----TGG 2284  
QY 381 MetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyGlyThrGly 400  
Db 2285 ATGGGG---CGCTGGCGATCGATCTCGCAGCGCAACCATCGGTGTTCGTGACCGCG 2341  
QY 401 AlaThrLeuTyAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIle 420  
Db 2342 TACCGCATCTGGCGCTCGCGCAATCTG-----CAGGATTCGCC 2380  
QY 421 AlaPro-----MetValLysGlyLeuGluThrAlaVal 432  
Db 2381 GCACCGCAGCGCGCGCTGCGTGGTGGTTCAGGACCGTGGGTGGAGAAACCGTCCG 2440  
QY 433 AsnAspLeuIleSerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGly 452

Db 2441 CTGGACCTGCTCAGCCCGATGGTGGCGCGCATCTGCTCAGCGCGCTCGCGGATATCAAC 2500

Qy 453 GlyPheThrHisAlaAspValThraValProSerThrIlePheThrSerProValPhe 472

Db 2501 GGTTCGGCATGACGACCTGGACCGCGTG--CAGTTGCAGTACCGCGCCCGCGGCTG 2557

Qy 473 ThrThrGlyThrSerValAspTyrAlaGluIleuAsnProSerIleIleValArgAlaGly 492

Db 2558 ACCAATGGCGAAAGCATCGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGT 2617

Qy 493 SerPhe-AspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly 512

Db 2618 ACCGTGGCG 2674

Qy 512 YIyeAsnTrpPheGln---GlySerGluProGlyGlyValThrThrGlyGlyThrValAl 531

Db 2675 CAAGCATGGACTGGTTTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGT 2734

Qy 531 aAlaSerAlaAspGlySerArgPheValTyrAlaProGlyAspProGlyGlnProValVa 551

Db 2735 CATTTGGTCCGATGCAAT 2788

Qy 551 lTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGl 571

Db 2789 GCGCAGTCCGACTTCG 2848

Qy 571 nIleArgSerAspArgValAsnProLysThrPheTyrAla-----LeuSerAsnGlyTh 589

Db 2849 GGTGTGACCATCGGTGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2908

Qy 589 rPheTyrArgSerThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeu---- 607

Db 2909 GCTGTACGAGACCGCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2962

Qy 608 -----ProSerSerGlyAlaVa 613

Db 2963 GCGCAGCG 3022

Qy 613 lGlyValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSe 633

Db 3023 AGCGTGTGTATCTGGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3053

Qy 633 rGlyLeuTyrHisSerThrAsnGlyLysSerTrpSerAlaIleThrGlyValSerSe 653

Db 3054 -GGTGTGTGCGTGGCAGGACGGT-----CGCTGCGAGTGTCTTCACAGCGCGGACA 3106

Qy 653 rAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheVa 673

Db 3107 AGCAGCTCGCTGGCATCGGCAAGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGT 3166

Qy 673 lValGlyThrIleGlyGlyValThrGlyAlaTyrArgSerAspCysGlyThrThrTr 693

Db 3167 GCGCGCGCGCGTGGAGCGGTGGATGGGTGTTCGCTCCGACGATGGCGCGCGCGCGCG 3226

Qy 693 pValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAs 713

Db 3227 GCAGCGCATCAACATGACGCGCACCGCTTCGGCGCG-----CGGTACAGCGTACCGGTGA 3283

Qy 713 pHisAlaAsnLeuArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAs 733

Db 3284 TCCGGCATTCGCGCGCGGTGTGTACTTCGCCACCGCGCGCGCGCGCGCGCGCGCGCG 3343

Qy 733 p 733

Db 3344 T 3344

## RESULT 10

US-10-395-241-13  
 ; Sequence 13, Application US/10395241  
 ; Publication No. US20040038367A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: YAOI, Katsuro  
 ; APPLICANT: MITSUISHI, Yasuhide

; TITLE OF INVENTION: NOVEL XYLOGLUCAN OLIGOSACCHARIDE-DEGRADING ENZYME, POLYNUCLEOTIDE  
 ; FILE OF INVENTION: ENCODING THE SAME, AND METHOD OF PREPARING THE ENZYME  
 ; FILE REFERENCE: Q73756  
 ; CURRENT APPLICATION NUMBER: US/10/395,241  
 ; CURRENT FILING DATE: 2003-03-25  
 ; PRIOR APPLICATION NUMBER: JP 2002-83433  
 ; PRIOR FILING DATE: 2002-03-25  
 ; NUMBER OF SEQ ID NOS: 19  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 13  
 ; LENGTH: 2367  
 ; TYPE: DNA  
 ; ORGANISM: Geotrichum sp. M128  
 ; FEATURE:  
 ; NAME/KEY: mat\_peptide  
 ; LOCATION: (1)..(2367)  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1)..(2367)  
 ; US-10-395-241-13

## Alignment Scores:

Pred. No.: 5,94e-100 Length: 2367  
 Score: 1117.00 Matches: 281  
 Percent Similarity: 50.37% Conservative: 124  
 Best Local Similarity: 34.95% Mismatches: 279  
 Query Match: 27.68% Indels: 120  
 DB: 13 Gaps: 31

US-09-917-376-3 (1-740) x US-10-395-241-13 (1-2367)

Qy 6 TyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn 25

Db 10 TACGAGTTCAGAGATGTCGCGATCGCGCGCGCGGGTACATTACCGGGATTGTCGCGCAC 69

Qy 26 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrpAsp 45

Db 70 CCAAGACCAAGGACCTGCTGTACGCGCGCACGACATTGGCGCGCGTACCGCTGGGAC 129

Qy 46 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr 65

Db 130 GCAGGCAGCTCAAGTGGATCCGCTCAACGACTTATTCGAGCGCGCAGCATGAACATT 189

Qy 66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 85

Db 190 ATGGCCACCGAGTGCATCGCTGCGACCCCAACACCCGACAGGCTGTACCTCGCGCAG 249

Qy 86 GlyMetTyrThr---AsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAsp 104

Db 250 GGGCGCTATGTCGGCGACGAGTGG-----GCGCGCTTCTATGTGTCGGAAGAC 297

Qy 105 GlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGly 124

Db 298 CGCGCCAGTCTTACAACTTACAGTCCGCTTCCCGTCCCGTCCCGACGACGATGGGA 357

Qy 125 ArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGly 144

Db 358 CGCAACAATGGCGAGCGCTCGCTCAACCGTTCACCTCGAACGAGGTCTGGATGGGT 417

Qy 145 AlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMet 164

Db 418 ACGCGTACA---GAGGGTATCTGGAAGAGATTTCGACCGCGCCCAAGACCTGGACAAAGTC 474

Qy 165 ThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGln 184

Db 475 ACGTCCATCCGCGACGCGTTC-----ACCAACGCTATCGGATACACG 516

Qy 185 SerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerSerLeuGlyGln 204

Db 517 TCG-----GTCAATTTTCGACCCC-----GAA 537

Qy 205 AlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSerArg 224

Db 224

Db 538 CGTAATGGACCATCTACGGAGCGGCGACTGCCCGCAGGGC---ATGTACGTACGCAC 594  
QY 225 AspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIle----- 241  
Db 595 GACGGCGGTGTCTCGTGGGAGCAGTGGCGGCGCCAGCCGTCAGCTCCAAAGGACC 654  
QY 242 -----ProHisGlyVal 246  
Db 655 ACGGGCGGTTCGGGACAAAGAGCCGCGTGCATCGCGCCGACGCCCAAGTGCCT 714  
QY 247 PheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAsp 266  
Db 715 CTCACCCCC-----AACTCTCTACGTGACTTACGGCGACATACCTGTGTCCATGGGCG 768  
QY 267 GlySerSerGlyAspValTrpIlePheSerValThrSerGlyThrTrpThrArgIle---- 285  
Db 769 GTCACGTTTCGGCGAAGTCTGGCGCAGAACCGACCTCGGGCGCTGGGACGACATTACT 828  
QY 286 -----SerProValProSerThrAspThrAlaAsnAspTyrPheGly 299  
Db 829 CCCCAGCGTGGCAACTCGTCCGCTGCCCGTACAAACACACGACGTTCCCTCGGGCGGA 888  
QY 300 TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIle 319  
Db 889 TTTTCGGTCTCAGGTTCAGCGCAGCGCAACCCCAACCGTCTGTGTGTATACC---CTC 945  
QY 320 SerTrpTrpProAspThrIle-----IlePheArgSerThrAspGlyGlyAlaThr 336  
Db 946 GACCGCGACCGGACCCCGCTCGACGACATCTACCTCTCAACCGATCGCGCGGACCC 1005  
QY 337 TrpThrArgIleTrpAspTrpThrSer-----TyrPro 347  
Db 1006 TGGAAAGGACGTACCCAGCTCTCGTCCCGCTCCAACTCGAAGTAACTGGGGCCACCCG 1065  
QY 348 AsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpLeuThrPheGlyVal 367  
Db 1066 ACTAACCGCGCGGTAC---AAGGACGCGACCGCTTCCTGGTGGCTCGACTTCAACAAC 1122  
QY 368 GlnPro-----AsnProProValProSerPro-----LysLeuGly 379  
Db 1123 GTTCCCGAGTGGGGGATACGGTGGCGCGACGGTACGCCCGGCTCACCAAGTTTGGC 1182  
QY 380 TrpMetAspGluAlaMetAlaIleAspProPheAsnSerArgMetLeuTyrGlyThr 399  
Db 1183 TGGTGGATGAGCGCTGTGTCTATCGATCGTTTCAACCCCGACGACCTGATGTACGCGAC 1242  
QY 400 GlyAlaThrLeuTyrAlaThrAsnAspLeuThrLys-----TrpAspSerGly 415  
Db 1243 GGGGCGACCATCTGGCGGACCGACACGCTCTCCCGTGTGAGAAAGACTGG----- 1293  
QY 416 GlyGlnIleHisIleAlaPro-----MetValLysGlyLeuGluThrAla 431  
Db 1294 -----GCGCGAGCTGGTACTTCCAGATCGAGTATCGAGAGATGCG 1338  
QY 432 ValAsnAspLeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeu 451  
Db 1339 ATCTGTCTCGCTCGCTCGCCCAAGAGCGCGCGCTCTCTGCGGCATCGTGCACATT 1398  
QY 452 GlyGlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProVal 471  
Db 1399 ACGCGCATGAACACGACGACGACCTTCAACAG---CCCCAGAAAGTGTGTGTGCGCCCGCAG 1455  
QY 472 PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAla 491  
Db 1456 TTCTCCAACTCGACAGCATCGACGCTCGGGCAACTTCCCAACGTTGTCTCGCGCC 1515  
QY 492 GlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly 511  
Db 1516 GGATCTCTCGGACACGAGTACGACACGCGGTGGCGCGGTGGCTACGCGACTGACGCG 1575  
QY 512 GlyLysAsnTrp-----PheGlnGlySerGluProGlyGlyValThrThr----- 526  
Db 1576 GGAGACGGGTGGACCATCTTCCCTACCTGCCCTCTGGCATGAACGCGGACCACTACCAG 1635

QY 527 GlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPro-----Gly 544  
Db 1636 GGCAGCACCATTCAGTTCGACGCGGCGGCGAGCATCGTGTGTCGACCAAGCTTGAC 1695  
QY 545 AspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGln 564  
Db 1696 GAGCAGGCGCTCGGACCGTGTACTCGCAGCATATGCGCAAGCGTGGTCT----- 1746  
QY 565 GlyValProAla-----AsnAlaGlnIleArgSerAspArgValAsn 578  
Db 1747 ---GTTCCCGCTGGCGACCTGAAGGCCGACACTGCCAATGTGTCTCGGACAGGTCCAG 1803  
QY 579 ProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyVal 598  
Db 1804 GATGGCACGTTCTACGCTACCGATGGCGCAAGTCTCTCTCGTCTCGACCGCGCGGAAG 1863  
QY 599 ThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHis 618  
Db 1864 TCGTATGCCCGCAAGGCGCGGACTTGTCACT-----GGCACATGCTCATGCGCTGCC 1917  
QY 619 AlaValProGlyLysGluGlyAspLeuTrpLeu---AlaAlaSerSerGlyLeuTyrHis 637  
Db 1918 GTGAACCCCTGGGTGGCGCGGACGCTGGGTGCTGTTCCGAGGGCGGTCTCTTCCAC 1977  
QY 638 SerThrAsnGlyGlySerSerTrpSerAlaIle---ThrGlyValSerSerAlaValAsn 656  
Db 1978 TCGACCGACTTTGGCGCTCTGTTCAGCGGGTAGGTACCCCAACGCGACCTCGTGAGC 2037  
QY 657 ValGlyPheGlyLysSer-----AlaProGlySerSerTyrProAlaValPhe 672  
Db 2038 GTCCGCGCCCCCAAGTCCAAGTCGACGCGCAAGAGCTAGCGCGCTCCCGGCTTTC 2097  
QY 673 ValValGlyThr-----IleGlyGlyValThrGlyAlaTyrArgSerAspCysGly 690  
Db 2098 ATCTGGGCGACGACGACAGGCTGGAAGCGACATCGGCTGTACCGTCCGACGACACGCG 2157  
QY 691 ThrThrTrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIle 710  
Db 2158 AGCAGCTGACGCGCTCAATGACGAGGACACAACTACTCTCGGGC---CCACCATGATC 2214  
QY 711 ThrGlyAspHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIleVal 730  
Db 2215 GAGGCGCGACCCCAAGGCTCTACGCGCGCTGTATCTAGGCACGAACGCGCGGTATCGTG 2274  
QY 731 TyrGlyAspIle 734  
Db 2275 TAGCGCGACCTT 2286

## RESULT 11

US-10-395-241-11  
; Sequence 11, Application US/10395241  
; Publication No. US20040038367A1  
; GENERAL INFORMATION:  
; APPLICANT: YAOI, Katsuro  
; APPLICANT: MITSUISHI, Yasushi  
; TITLE OF INVENTION: NOVEL XYLOGLUCAN OLIGOSACCHARIDE-DEGRADING ENZYME, POLYNUCLEOTIDE  
; FILE OF INVENTION: ENCODING THE SAME, AND METHOD OF PREPARING THE ENZYME  
; FILE REFERENCE: Q73756  
; CURRENT APPLICATION NUMBER: US/10/395,241  
; CURRENT FILING DATE: 2003-03-25  
; PRIOR APPLICATION NUMBER: JP 2002-83433  
; PRIOR FILING DATE: 2002-03-25  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: Patentin version 3.2  
; SEQ ID NO 11  
; LENGTH: 2646  
; TYPE: DNA  
; ORGANISM: Geotrichum sp. M128  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (120)..(2558)  
US-10-395-241-11







QY 619 AlaValProGlyLysGluGlyAspLeuTrpLeu---AlaAlaSerSerGlyLeuTyrHis 637  
 DB 2106 GTGAACCCCTGGGTGGCGGCGACGCTCTGGTGGCTGTTCCTCCAGGCGGCTCTCTTCCAC 2165  
 QY 638 SerThrAnGlySerSerTrpSerAlaIle---ThrGlyValSerSerAlaValAsn 656  
 DB 2166 TGAACCGAGTGGCGGCTCTGGTTCAGAGGGTAGGTACCGCAACGCGACCTCGTGAGC 2225  
 QY 657 ValGlyPheGlyLysSer-----AlaProGlySerSerTyrProAlaValPhe 672  
 DB 2226 GTCCGGGCCCCAAGTCCAAGTCGACGCGCAAGAGGTAGCGCGCTCCCGGCTCTTC 2285  
 QY 673 ValValGlyThr-----IleGlyGlyValThrGlyAlaTyrArgSerAspCysGly 690  
 DB 2286 ATCTGGGGCAGCGCAAGCCTCGAAGCGACATCGGCCTGTACCGCTCCGACGACAACGCG 2345  
 QY 691 ThrThrTrpValLeuIleAsnAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIle 710  
 DB 2346 AGCAGCTGGAGCGCGCTCAATGACGAGGAGCAACTACTCGGCG---CCCAACCATGATC 2402  
 QY 711 ThrGlyAspHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIleVal 730  
 DB 2403 GAGCGCGACCCCAAGTCTACGGGCGGTGTATAGGACGACGCGCGGTATCGTG 2462  
 QY 731 TyrGlyAspIle 734  
 DB 2463 TACGCGGACCTT 2474

## RESULT 12

US-10-395-241-17  
 ; Sequence 17, Application US/10395241  
 ; Publication No. US20040038367A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: YAOI, Katuro  
 ; APPLICANT: MITSUISHI, Yasushi  
 ; TITLE OF INVENTION: NOVEL XYLOGLUCAN OLIGOSACCHARIDE-DEGRADING ENZYME, POLYNUCLEOTIDE  
 ; TITLE OF INVENTION: ENCODING THE SAME, AND METHOD OF PREPARING THE ENZYME  
 ; FILE REFERENCE: Q73756  
 ; CURRENT APPLICATION NUMBER: US/10/395,241  
 ; PRIOR FILING DATE: 2003-03-25  
 ; PRIOR APPLICATION NUMBER: JP 2002-83433  
 ; PRIOR FILING DATE: 2002-03-25  
 ; NUMBER OF SEQ ID NOS: 19  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 17  
 ; LENGTH: 2481  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Xyloglucan Oligosaccharide-Degrading Enzyme with Histidine Tag  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1)..(2481)  
 US-10-395-241-17

## Alignment Scores:

Pred. No.: 1,258-99 Length: 2481  
 Score: 1114.00 Matches: 281  
 Percent Similarity: 50.25% Conservative: 123  
 Best Local Similarity: 34.95% Mismatches: 280  
 Query Match: 27.60% Indels: 120  
 DB: 13 Gaps: 31

US-09-917-376-3 (1-740) x US-10-395-241-17 (1-2481)

QY 6 TyrThrTrpSerAsnValAlaIleGlyGlyGlyGlyPheValAspGlyIleValPheAsn 25  
 DB 13 TACGAGTTCAGAAATGTCGGATCGCGGCGGGGTACATTACCGGATTTGTCGGCAC 72  
 QY 26 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyMetTyrArgTrpAsp 45  
 DB 73 CCAAGACCAAGGACCTGTGTACGCGCGCACGGACATTGGCGGCGGTACCGCTGGGAC 132

QY 46 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr 65  
 DB 133 GCAGGCACCTCCAAAGTGGATCCGCTCAACGACTTTATCGAGCGCGACGACATGAACATT 192  
 QY 66 AsnGlyValValSerIleAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 85  
 DB 193 ATGGGCGACCGAGTTCGATCGCTGGACCCCAACACCCGACAGGTGTACCTCGCGCAG 252  
 QY 86 GlyMetTyrThr---AsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAsp 104  
 DB 253 GGGCGGTATCTCGGCGACGAGTGG-----GCGCGTCTCTATGTCTCGAAGAC 300  
 QY 105 GlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGly 124  
 DB 301 CGCGGCGAGTCTGTTTACAATCTACGAGTCCGCGTTCGCGATGGCGGCGCAACGACATGGA 360  
 QY 125 ArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGly 144  
 DB 361 CGCAACAATGGGAGCGCTCGCTGTCAACCCGTTCAACTCGAACGAGGTCTGGATGGT 420  
 QY 145 AlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMet 164  
 DB 421 AGCGTACA---GAGGGTATCTGGAAGAGTTCGACCGCGCCAGACCTCGACAAACGTC 477  
 QY 165 ThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGln 184  
 DB 478 AGTCCATCCGCGACGCGTTC-----ACCAACGCTATCGGATACACG 519  
 QY 185 SerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerSerLeuGlyGln 204  
 DB 520 TCG-----GTCAATTTTCGACCCC-----GAA 540  
 QY 205 AlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSerArg 224  
 DB 541 CGTAATGGCACCATTCTACCGGAGCGGACTGCCCGCGAGGCG---ATGATCGTCACGAC 597  
 QY 225 AspGlyGlyAlaThrTrpGlnAlaProGlyAlaProThrGlyPheIle-----241  
 DB 598 GACGGGGTGTCTCGTGGGAGCAGTGGCGGCGCAGCGCTCCAGCTGGCTCAACAGGACC 657  
 QY 242 -----ProHisLysGlyVal 246  
 DB 658 ACGGCGCGCTTCCCGGACAAAGCCGCGTTCGATCGCGCGCGAGCCATCAAAAGTCCGCT 717  
 QY 247 PheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAsp 266  
 DB 718 CTCACCCCC---AATTCCTCTACGTGACTTACGCGGACTACCTCGTCCATGGGCG 771  
 QY 267 GlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrpThrArgIle---285  
 DB 772 CTCACGTTCCGCAAGTCTGGCGCCAGAACCGCACCTCGGCGCGCTGGGCGGACATTACT 831  
 QY 286 -----SerProValProSerThrAspThrAlaAsnAspTyrPheGly 299  
 DB 832 CCGCGCGTCGGCAACTCGTCGCTGCCCGGTACAAACACGAGAGTTCCTCGCGGCGG 891  
 QY 300 TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIle 319  
 DB 892 TTTTGGGTCTCAGCGTCGACGCGACCAACCCCAACCGTCTCGTCTGTCATCAC---CTC 948  
 QY 320 SerTrpTrpProAspThrIle-----IlePheArgSerThrAspGlyGlyAlaThr 336  
 DB 949 GACCGGCGCCCGGACCCCGCTCGACAGCACTCTCTCAACGATGCGCGGCGGAC 1008  
 QY 337 TrpThrArgIleTrpAspTrpThrSer-----TyrPro 347  
 DB 1009 TGAAGAGCGTCACCCAGCTCTCGTCCCGTCCCAACTCGAAGGTAACGTGGGCGCACCCG 1068  
 QY 348 AsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpLeuThrPheGlyVal 367  
 DB 1069 ACTAAGCGCGCGGTAC---AAGGACGCGACGCGCTGTTCGTCGTTCGACTTCAACAAC 1125



QY 85 ----- 85  
Db 662 ACCCGTAGTCGCGAACCGAATAGCGCTCTGCTAGTGTTCACATACGGCGCTAC 721  
QY 86 GlyMetThrAsnSerTrpAsp----- 93  
Db 722 GGCATGTATACGAACAGCTGGTC-TGTGATGCTCTCAGATCTAGACCTATGATGGAGC 780  
QY 93 ----- 93  
Db 781 GCCGTACATATGCTTGTGACCCAGACACTACAGAGAGTCTAGATCTGGATACTAACTGC 840  
QY 94 -----ProAsnAspGlyAlaIleLeuArgSerSerAspGln 105  
Db 841 CTGACATTGGCCATATATAGGATCCGAGTAAATGAGGCCATCAITTCGCTCGCAGACCGC 900  
QY 105 ----- 105  
Db 901 GACTGTAAACGGTATATATCCCTAGCTCAITTCCTCGGTAGTAAGCGAGAGCTGGCG 960  
QY 106 GlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArg 125  
Db 961 GCGCAACAGTGTCTCTCAACCAACTTGCCCTTCAAAGTCGGGGGTAACATGCCAGAGCG 1020  
QY 125 ----- 125  
Db 1021 CGCGCTGCACAGGAAGTGTGTGAACGGGAAGTTTCAGCCCCCATTTGTACGGTCTCTGCG 1080  
QY 126 GlyMetGlyGluArgLeuAlaValAspProAsnAspAsnIleLeuTyrPheGlyAla 145  
Db 1081 GGAGCGGAGAGCGTCTGCTGTGCATCGGCCAACTCCCAACATCATCTACTTTGTGTCT 1140  
QY 146 Pro----- 146  
Db 1141 CCTCGCCTCTCGCAGCCGACAGCTAGCGCGTTGAGTGTCTAGTAGTAAACCAACA 1200  
QY 147 ---SerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMet--- 164  
Db 1201 CGCTCAGGAACGGCTCTGGAAGTCTACGACGGCGGCGTGACCTTTTCCAAAGGTCTCG 1260  
QY 165 ---Thr 165  
Db 1261 GCGAGTCTTTGCCGAGACCTTCAGATGCTCGCGCGCCTGGAAGGTTCCAGAGC 1320  
QY 166 AsnPheProAspValGlyThrTrpIleAlaAsnProThrAspThrThrGlyTyrGln--- 184  
Db 1321 TCGTTACGGCAACTGGGACGCTACATCCAGACCCGAGTGATCCAAACGGCTACACAGC 1380  
QY 185 -----Ser 185  
Db 1381 AGCAAGTGCCTGTGACCTCGATGTAGGTCTGGGCTCACTAAGGTGGCGATGTTGTGC 1440  
QY 186 AspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerLeuGlyGlnAla 205  
Db 1441 GACAAAGCAAGGACTCATGTGGGTACGTTCGATCAACAGCAGCAGCAGCGGGGAGCC 1500  
QY 205 ----- 205  
Db 1501 CTGTTCTGTTCTAGTACACCAATGCAAGCTGAGTGTGCTGCTGGCGCCCTCGC 1560  
QY 206 SerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrp----- 222  
Db 1561 ACGTCTGTATCTTTGTTGGCAGCGCTGATACATCATCTGCTTCAGTCTATGTGAGCAGC 1620  
QY 223 -----SerArg 224  
Db 1621 TGCAGAGCATAGAAACAAACCGTCGCGACTATTTGTAGTGACGAAGTCAGATACACTCGTGC 1680  
QY 225 AspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIleProHis--- 243  
Db 1681 AATGCCGCTCCACGTGGAGTGTGTACCGGGGACCGAGGGAATATCTTCTCTCACAG 1740  
QY 244 -----LysGlyValPhe 247

Db 1741 TTACGCCCGAGGTGCACCTCAGACATGGCCCCGTGCTCCCTTTATGAAAGAGTGTTC 1800  
QY 248 -----AspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyPro 264  
Db 1801 GCGAAACTGCAGCCAGCAGAGAAGGCTTGTATCTGACCTATTCCGATGCGACAGGCGCG 1860  
QY 264 ----- 264  
Db 1861 CGCTTTGAGCTCGGTCTCTCTCCGGAACATAGACTGGATAAGGCTACCGTCTCCCGC 1920  
QY 265 TyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrpThrArg 284  
Db 1921 TATGATGGCACACTTGGCTCAGTGTGGAGGTACACATTGCGAGGGGAACCTTGGAAAGAC 1980  
QY 285 IleSerProValProSerThrAspThr----- 293  
Db 1981 ATACTACCGTGTGAACCGAGTCACACTCCATGCTGTAAAGTCCCTCTGAACCTTCTG 2040  
QY 294 -----AlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAsp--- 306  
Db 2041 ATCACCCCTGTCTCTGGATCAGATCTATCTTTGGCTTTGGCGCCTTGGCCTCGATTG 2100  
QY 307 -----Arg 307  
Db 2101 TAGTGGGACAGAGACTAGTCTAGATATGAACCGAAACCGCCGGAACCGAGCTAAAC 2160  
QY 308 GlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrpProAspThr----- 325  
Db 2161 CAAAGCCAGGAACCTTGTGTGTCTTCTTGAACCTTTGTGGCCAGATGCTCAGCTG 2220  
QY 326 -----IleIle 327  
Db 2221 GTTTTCGGTCTTGGGAACAACAAGAAACTTGAACACACCGCTCTACGAGTGCAC 2280  
QY 328 PheArgSerThrAspGlyGlyAlaThrTrpThrArgIleTrpAspTrpThrSerTyrPro 347  
Db 2281 TTTTCGGTCCAGCGACTCTCGGACACATCGAGCCCGATCTGGCGCTGGCGAGTATCCG 2340  
QY 348 AsnArgSer---LeuArgTyrValLeuAspIleSerAlaGluPro----- 361  
Db 2341 AAAGCCAGCTGGCTGAGACC-CTGTTGTACCTCGGGCTAGACCCCGCACCCGCTCGATAGG 2399  
QY 362 -----Trp 362  
Db 2400 CACTGAGACCTATTACTACAGCATCTCAGTGAGTCACTCTTAACGATCCGATCGAATGG 2459  
QY 363 LeuThrPhe-----GlyValGlnProAsn----- 370  
Db 2460 CTGACTCTGGATAATGATGTCTGAGTCACTCAGTGAGAATTGCTAGGCTAGCTTACC 2519  
QY 371 -----ProProValPro----- 374  
Db 2520 GACTGACTCTGTTCAAGACTCCCAAGCACCGTGGATCAAGAACACTTTTATCGATGTG 2579  
QY 374 ----- 374  
Db 2580 ATGACTGACACGAAGTCTTGAGGGTTTCGTGGCACCTAGTCTTGTGTAATAGCTACAC 2639  
QY 375 -----SerPro-----LysLeuGlyTrpMetAspGluAlaMet 385  
Db 2640 TCGAGCGAGTCCACCGTCCGATGCTCTCAACAGCGCTCGGCTGGATGATTGAGTCTCTC 2699  
QY 386 Ala----- 386  
Db 2700 GGCTCGCTCAGTGGCAGGCTACAGAGTAGTTCGGGAGCCGACCTACTAACTCAGAGAG 2759  
QY 387 ---IleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAla 405  
Db 2760 CAGATTGACCCCAACCCAGACGAACCACTCGCTCTACGGCACCGGAATGACAACTTTTGGC 2819  
QY 406 Thr----- 406



```

; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA-034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30151
; LENGTH: 7407
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-10-282-122A-30151

Alignment Scores:
Pred. No.: 1,278-12 Length: 7407
Score: 238.00 Matches: 205
Percent Similarity: 32.59% Conservative: 88
Best Local Similarity: 22.80% Mismatches: 314
Query Match: 5.90% Indels: 292
DB: 13 Gaps: 50

US-09-917-376-3 (1-740) x US-10-282-122A-30151 (1-7407)
QY 11 ValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGly 30
Db 3250 GTCAACCTGAGCAATGGCAGCAGCCTCAGCGC-----ACCGCCGAGCGCGC 3297
QY 31 IleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTyrPheAlaAlaAsnGlyArg 50
Db 3298 AGCAGCGGTGATTCTCACCAGC-----GGCAACGGCAAT 3330
QY 51 TrpIleProLeuLeuAspTrpValGlyTyrPheAsnThrGlyTyr----- 65
Db 3331 CCGATCGCGAGGTATCACCGCAGCGCAGCGCACTGACCTACACCCCGTCCACGCCG 3390
QY 66 -----AsnGly---ValValSerIleAlaAlaAspProIleAsnThrLysValTyr 82
Db 3391 ATCGCAACGGTACTGTGTGTCAAGGTGTGGCCAGGACCTCCGGTAAACAGCAGCCCG 3450
QY 83 AlaAlaValGlyMetTyrThrAsnSer-----TrpAspProAsnAsp 96
Db 3451 CCGGGCGAGCGTACCGTCGATTCAGCGCGCGCGCGCGCGGTGATCAACCGAGCAAC 3510
QY 97 GlyAlaIleLeuArg---SerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuPro 115
Db 3511 GCGGTGTCATCAGCGGCAACCGCGGCGCGGTGCGGACCGGTGACCCCTCACC----- 3561
QY 116 PheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspPro 135
Db 3562 ----GATCGCGCGGCAACCGGATAGGCGAGGTACCGCGCGC----- 3600
QY 136 AsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTyrArgSerThr 155
```

```

Db 3601 -----GGCAGCGCAACTGGAGCTTCCAG 3624
QY 156 AspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAla 175
Db 3625 CCG-----GGCAGCGCGCTGGCC 3642
QY 176 Asn-----ProThrAspThrThrGlyTyrGlnSer 185
Db 3643 AACGGCACGGTGTGTCGCGCACCGCCAGCCGACCGGCAATACCGGCGCGCAGGCC 3702
QY 186 -----AspIleGlnGlyValTyrValAlaPheAspLysSerSerSer 201
Db 3703 GCCACCAACGGTGGAGCGGTGGCGCGCGCGGTGATCGATCCGAGCAACGGCACG 3762
QY 202 Leu-----GlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsn---Asn 218
Db 3763 ACCATCAGCGGACCGCGGAGCGCGGAGGTGATCTCACCGAGCGGCAACGGCAAC 3822
QY 219 ProVal---PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaPro 237
Db 3823 CCGATCGGCAAAACCAACCGCGCAGCGCAACTGGAGCTTCACGCCCGCGCACGCCG 3882
QY 238 -----ThrGly--- 239
Db 3883 CTGGCCAAACGGCACGGTGTCAACCGCGTGGCCCGAGGACCCCTGGGGCAATACCGGCCG 3942
QY 240 -----PheIleProHisLysGlyValPheAspProValAsn 251
Db 3943 CAGGGGAGCACATACCGTGGAGCGGTGGCGCGGCAACAGCGCTGTGGTCAATCCGCAAC 4002
QY 252 HisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyr-----AspGlySer 268
Db 4003 GGCAACCTGCTCAACGGTACCGCGCAGCGCGGAGCACCGTGGACCGTTCACCGAGCGCAAC 4062
QY 269 SerGlyAspValTrpLysPheSerValThr---SerGlyThrTrp-----ThrArgIle 285
Db 4063 GGCAACCCGATCGCGCAGACCAACCGCGATGGCAGCGGCAACTGGAGCTTCACGCCCGCG 4122
QY 286 SerProValProSer-----ThrAspThrAlaAsnAspTyrPheGlyTyrSer 301
Db 4123 TCGAACTGCCCAACCGCACCGCGTGTCAACGTACCGCGCGGAGCGCGCGCGGCAATACC 4182
QY 302 GlyLeu-----ThrIleAspArgGlnHisProAsnThrIleMetVal----- 315
Db 4183 AGCTTCCCGCTACCAACGAGCGGTGATTCCTCGTCGCTCGATCCCGCAGGTGGATCCG 4242
QY 316 -----AlaThrGlnIleSerTrpTrpProAsp-----ThrIleIlePheArgSer 330
Db 4243 AGCAACGGTTCGGTGTATCAGCGCGCACCGCGCGCAACCACTCATCATC----- 4296
QY 331 ThrAspGlyGly-----AlaThrTrpThrArgIleTrpAsp 342
Db 4297 ACCGATGGCAACGGCAACCGGATTCGCGAGTCAACCGCGCGCGCGCAACTGGTCC 4356
QY 343 TrpThr-----SerTyrProAsnArgSerLeuArgTyrVal 354
Db 4357 TTCCTCCAGGATCCCGTTCGCGGATGGACCGGTGTCAACGTGGTGGCGCGCGCGCA 4416
QY 355 LeuAspIleSerAlaGluProTrpLeu-----ThrPheGlyValGlnProAsnProPro 372
Db 4417 AGCAATGTGCACAGTACCGCGCGGTGTATCATGTGTGTGGTGGCGCGCGCGCGCGCG 4476
QY 373 ValProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSer 392
Db 4477 GTG-----ATCGATCCGAGCAACCGC 4497
QY 393 AspArgMetLeuTyrGlyThr-----GlyAlaThrLeuTyrAlaThrAsnAspLeu 409
Db 4498 ACGGAG---ATAAGCGGTACCGCGGAGCGCGCGCGGTGATCTCTCACC----- 4545
QY 410 ThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluGlu 429
```

Db 4546 -----GATGGCGGGGCAAC-----CCGATCGCGCCAGGCCACCGCCGAC 4584  
Qy 430 ThrAlaValAsnAspLeuLeuSerPro-----ProSerGlyAlaProLeuLeu 445  
Db 4595 GCGCGGCAACTGGACGTTTCAACCCGAGCAGCCCGCTGGCCACGCGCACCGTGTCAAC 4644  
Qy 446 SerAlaLeuGlyAspLeuGlyPheThrHis-----AlaAspVal 459  
Db 4645 GCGGTGGCCAGGACCGCGCGGCAATACCAAGCGGTCCGGCCAGCGTCAACCGTGTGATCC 4704  
Qy 460 ThrAlaValProSerThrThrPhe-----ThrSerProValPheThrThrGlyThrSer 477  
Db 4705 ATCGCCCGCGCGCGCGGTGATCAATCCGAGCAATGAGTGTGATCATCAGCGGTACG--- 4761  
Qy 478 ValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSer 497  
Db 4762 -----GGGAAGCGGGGCGCGGTGATCTCCAGCGGCAACGCGCACCGGATC 4812  
Qy 498 SerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPhe--- 516  
Db 4813 GGCAG-----GTACCGCGGACGCGCGGCAACTGGAGCTTC 4851  
Qy 516 ----- 516  
Db 4852 AGCCCGGACGCGCGCTGGCCACGCGTGGTGTGATCAATGCGTGGCGCCAGGACGCGCC 4911  
Qy 517 ---GlnGlySerGluProGlyGlyValThr----- 525  
Db 4912 GGCACACAGCAGTCCACACGCGGCCACCGTCACTGCTGGCGGCCAGCGCCCGGTG 4971  
Qy 526 -----ThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrp 541  
Db 4972 ATCGATCCGACCAACGAGTGGTGTGATCGCGGTACCGCGAGGTGGGCCAGGTGATC 5031  
Qy 542 AlaProGlyAspProGlyGlnProVal-----ValTyrAlaValGlyPheGlyAsnSer 559  
Db 5032 CTCACCGACGCGCAACCGGCAACCGGCGGCGGAGTCAACCGCGGATGGCGCGGCAAC--- 5088  
Qy 560 TrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnPro 579  
Db 5089 TGGAGCTTCACGCGCGGACGCG--- 5112  
Qy 580 LysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArg----- 592  
Db 5113 -----CTGTCAATGGACGCGTGTGATCGGTGGTGGCCAGGACGCTGCC 5157  
Qy 593 ---SerThrAspGlyVal-----ThrPheGlnProValAlaAlaGlyLeuPro--- 608  
Db 5158 GGCACACCGCGCGCGGTGATCGACACCGGTGGACGCGGTGGCGCGCCAGCGCGGTG 5217  
Qy 609 -----SerSerGlyAlaValGlyValMetPheHisAlaValProGlyLysGluGlyAsp 626  
Db 5218 ATCGACCGGACCAACGCTGTCGAATCAAGCGGACCGCGGCAACCGCGGTGCGGTGATC 5277  
Qy 627 LeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSer---ThrAsnGlyGlySerSerTrp 645  
Db 5278 CTCACCGATGTCATGGCAATCCGATCGCGCAGACCCCTTGGCGGACGCGGCGGCACTGG 5337  
Qy 646 SerAlaIleThrGlyVal-----SerSerAlaValAsnValGlyPheGlyLys 661  
Db 5338 AGCTTCACGCGCGGACGCGCTGGCCACGCGGACGCGGTGGTCAATCGCGTGGCCAGGAC 5397  
Qy 662 SerAlaProGlySerSerTyrProAlaValPheValValGlyThrIleGlyGlyValThr 681  
Db 5398 CCGGCGCGCAATACCGCGCGCGGCGGCGGACCGGCGGCGGCGGCGGCGGCGGCGGCG 5457  
Qy 682 GlyAlaTyrArgSerAspAspCysGlyThrTrpValLeuLeuAsnAspAspGlnHis 701  
Db 5458 -----CCGCTGATCAATCCGAGCAAC--- 5478  
Qy 702 GlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgValTyr 721  
Db 5479 -----GGCAGCGTGTATCACCGGCAAC---GCCAGGTGCGGCGGCGGCGGCGG 5520

Qy 722 IleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyGlyAlaProSerGly 740  
Db 5521 ATCTCTACCGATGGCAAGCGCAACCGATCGCGGAGACACCGCCGCGCAGTGGT 5577

## RESULT 15

US-10-246-330-3  
; Sequence 3, Application US/10246330  
; Publication No. US20030166030A1  
; GENERAL INFORMATION:  
; APPLICANT: O'Toole, George A.  
; APPLICANT: Mah, Thien-Fah  
; TITLE OF INVENTION: METHODS TO STUDY AND MECHANISMS OF  
; FILE OF INVENTION: BIOFILM-SPECIFIC ANTIBIOTIC RESISTANCE  
; FILE REFERENCE: 14537-002001  
; CURRENT APPLICATION NUMBER: US/10/246,330  
; CURRENT FILING DATE: 2002-09-18  
; PRIOR APPLICATION NUMBER: US 60/323,241  
; PRIOR FILING DATE: 2001-09-18  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 7407  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1).....(7404)  
US-10-246-330-3

Alignment Scores:  
Pred. No.: 1,27e-12 Length: 7407  
Score: 238.00 Matches: 205  
Percent Similarity: 32.53% Conservative: 88  
Best Local Similarity: 22.80% Mismatches: 314  
Query Match: 5.90% Indels: 292  
DB: 15 Gaps: 50

US-09-917-376-3 (1-740) x US-10-246-330-3 (1-7407)

Qy 11 ValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGly 30  
Db 3250 GTCAACCTGAGCAATGCGCAGCAGCTCAGCGGC-----ACCGCGGAGCGCGC 3297  
Qy 31 IleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArg 50  
Db 3298 AGCAGCGGTATTCTCCAGC-----GGCAACGGCAAT 3330  
Qy 51 TrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr----- 65  
Db 3331 CCGATCGCGAGGTCAACCGCCAGCGGCGGCACTGACCTACACCGCTCCAGCGCG 3390  
Qy 66 -----AsnGly---ValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrp 82  
Db 3391 ATCGCCCAACGGTACTGTGTCAACGTGGTGCCAGGAGCGCTCCGTAACAGCAGCGCG 3450  
Qy 83 AlaAlaValGlyMetTyrThrAsnSer-----TrpAspProAsnAsp 96  
Db 3451 CCGGCGCGGTGACCGTCAATTCAGCGCGCGCGCGCGGTGATCAACCGCGGCAAC 3510  
Qy 97 GlyAlaIleLeuArg---SerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuPro 115  
Db 3511 GCGTGTGATCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 3561  
Qy 116 PheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspPro 135  
Db 3562 ---GATCGCGCGGCAACCGGATAGGCGAGGTCAACCGCGGAC----- 3600  
Qy 136 AsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThr 155  
Db 3601 -----GGCAGCGGCACTGGAGCTTCACG 3624  
Qy 156 AspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAla 175

Db 3625 CCG-----||||| :|||  
Qy 176 Asn-----ProThrAspThrThrGlyTyrGlnSer 185  
Db 3643 AACGGCAGGTGATCGTCCGACCGCCACCGACCGCGCAATACCGCGCCGAGGCC 3702  
Qy 186 -----AspIleGlnGlyValValTrpValAlaPheAspLysSerSerSer 201  
Db 3703 GCCACACGGTGGACCGGTGGCGCGCGCGCGGTGATCGATCCGAGCAACGGCACG 3762  
Qy 202 Leu-----GlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsn---Asn 218  
Db 3763 ACCATCAGCGGACCGCGGAGCGCGGGGCCAAGGTGATCTCACCGACGGCAACGGCAAC 3822  
Qy 219 ProVal---PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaPro 237  
Db 3823 CCGATCGCGGAACACACCGCGGACGGCAGCGCACTGGAGTTTCACGCCCGGACGCGG 3882  
Qy 238 -----ThrGly--- 239  
Db 3883 CTGGCCAAACGGCACGGTGGTCAACCGCGTGGCCAGGACCCTCGGGCAATACCGGCGCG 3942  
Qy 240 -----PheIleProHisLysGlyValPheAspProValAsn 251  
Db 3943 CAGGGCAGCACTACCGTGGACCGCGGTGGCGCGGAACACGCGCTGTGTCAATCCGAGCAAC 4002  
Qy 252 HisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyr-----AspGlySer 268  
Db 4003 GGCAACCTGTCAACGGTACCGCGGAGCGGGGACGACCGGTGACCTTGACCGACGGCAAC 4062  
Qy 269 SerGlyAspValTrpLysPheSerValThr---SerGlyThrTrp-----ThrArgIle 285  
Db 4063 GGCAACCGGATCGGCGCAGACACCGCCGATGSCAGCGCACTGGAGTTTCACGCGCGCGC 4122  
Qy 286 SerProValProSer-----ThrAspThrAlaAsnAspTyrPheGlyTyrSer 301  
Db 4123 TCGCAACTGCGCAACGGCACCGTGTGTCAACGTGTACCGGAGCGACGCGCGCGCAATACC 4182  
Qy 302 GlyLeu-----ThrIleAspArgGlnHisProAsnThrIleMetVal----- 315  
Db 4183 AGCCTTCGGGTACACAGCGGTGATTCCTCGTGGCGGTGATCCCGAGGTGATCCG 4242  
Qy 316 -----AlaThrGlnIleSerTrpTrpProAsp-----ThrIleIlePheArgSer 330  
Db 4243 AGCAACGGTTCGGTGTATCAGCGGCACCGCGGAGCGCGCAACACCATCATCATC----- 4296  
Qy 331 ThrAspGlyGly-----AlaThrTrpTrpArgIleTrpAsp 342  
Db 4297 ACCGATGGCAACGGCAACCCGATTGGCCAGGTTCACCGCGGACGCGGCAACTGTGTCC 4356  
Qy 343 TrpThr-----SerTyrProAsnArgSerLeuArgTyrVal 354  
Db 4357 TTCCTCAGGATCCCGCTCCCGATGGCAGGTGTCACCGTGGTGGCGCGGAGCCCA 4416  
Qy 355 LeuAspIleSerAlaGluProTrpLeu-----ThrPheGlyValGlnProAsnPro 372  
Db 4417 AGCAATGTCGACGTGCGCGCGGTGATCATCTGGATGGGTGGCGCGCGCGCGCG 4476  
Qy 373 ValProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSer 392  
Db 4477 GTG-----ATCGATCCGAGCAACGGC 4497  
Qy 393 AspArgMetLeuTyrGlyThr-----GlyAlaThrLeuTyrAlaThrAsnAspLeu 409  
Db 4498 ACCGAG---ATAAGCGGTACCGCGGAGCGCGCGCGGTGATCCTCACC----- 4545  
Qy 410 ThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGlu 429  
Db 4546 -----GATGGCGCGGGAAC-----CCGATCGGCGGAGCCACCGCCGAC 4584  
Qy 430 ThrAlaValAsnAspLeuIleSerPro-----ProSerGlyAlaProLeuIle 445  
Db :||| :|||

Db 4585 GGCAGGGCAACTGGAGCTTACCCCGAGCACCCTCGGTGCCCAACGGCAGCGTATCAAC 4644  
Qy 446 SerAlaLeuGlyAspLeuGlyGlyPheThrHis-----AlaAspVal 459  
Db 4645 GCGGTGGCCAGGACCGCGCGGCAATACACAGCGTCCGSCCAGCGTCACCGTCGATGCC 4704  
Qy 460 ThrAlaValProSerThrIlePhe-----ThrSerProValPheThrThrGlyThrSer 477  
Db 4705 ATCGCGCCCGCGCGGTGATCAATCCGAGCAATAGTGTCTCATCAGCGGTACG--- 4761  
Qy 478 ValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSer 497  
Db 4762 -----GCCGAAGCGCGGCGCACGCTGATCTCACCGACGGCAACGGCACCCGATC 4812  
Qy 498 SerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPhe--- 516  
Db 4813 GGCCAG-----GTCACCGCGGACGGCAGCGCACTGGAGGTTC 4851  
Qy 516 ----- 516  
Db 4852 ACGCCCGGCACGCGCTGGCCAAACGGCTCGGTGATCAATCGCTGGCCAGGACGCGGCC 4911  
Qy 517 ---GlnGlySerGluProGlyGlyValThr----- 525  
Db 4912 GGCAACAAACAGCAGTCCACACGCGCCCGTCTGCTGCTGGCGCCAGCAGCCCGGTG 4971  
Qy 526 -----ThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrp 541  
Db 4972 ATCGATCCGAGCAACGGTAGCTGTCGCGGTACCGCGGAGGTGGTGGCCAGGTGATC 5031  
Qy 542 AlaProGlyAspProGlyGlnProVal-----ValTyrAlaValGlyPheGlyAsnSer 559  
Db 5032 CTCACCGAGCGCAACGGCAACCGATCGCGCAGGTACCGCGCATGGCAGCGCAAC--- 5088  
Qy 560 TrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerArgValAsnPro 579  
Db 5089 TGGAGCTTCACCGCGGACGCGC----- 5112  
Qy 580 LysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArg----- 592  
Db 5113 -----CTGTCCAATGGCAGCGTGTCAATCGGTGGCCAGGAGGTGCC 5157  
Qy 593 ---SerThrAspGlyGlyVal-----ThrPheGlnProValAlaAlaGlyLeuPro--- 608  
Db 5158 GGCAACACACGCGCGCGGTCTACAGCACCGGTGGACGCGGTGGCCCGCGCCACCGCGGTG 5217  
Qy 609 -----SerSerGlyAlaValGlyValMetPheHisAlaValProGlyLysGluGlyAsp 626  
Db 5218 ATCGACCGCGAGCAACGGTGTCTGAACCTCAGCGGCACCGCGCAACCCGCGGTCCGGTGCATC 5277  
Qy 627 LeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSer---ThrAsnGlyGlySerSerTrp 645  
Db 5278 CTCACCGATGGCAATGGCAATCCGATCGCGCAGACCTTGTCCCGACGCGCAGCACTGG 5337  
Qy 646 SerAlaIleThrGlyVal-----SerSerAlaValAsnValGlyPheGlyLys 661  
Db 5338 AGCTTCACCGCGGACCGCGGTGGCCCAACGGCAGCGGTGTCTCAATGCCGTGGCCAGGAC 5397  
Qy 662 SerAlaProGlySerSerTyrProAlaValPheValGlyThrIleGlyGlyValThr 681  
Db 5398 CCGGCGCGCAATACCAACGCGCGCGCGCAGCACACCGCGGTGGACACCGGTGGTCCGCGCACG 5457  
Qy 682 GlyAlaTyrGlySerAspAspCysGlyThrTrpValLeuIleAsnAspAspGlnHis 701  
Db 5458 -----CCGGTGATCAATCCAGCAAC--- 5478  
Qy 702 GlnTyrAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgArgValTyr 721  
Db 5479 -----GGCAGCGTGTATCAGCGGCACC---GCCAGGTGGCGGCAAGGTG 5520  
Qy 722 IleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyGlyAlaProSerGly 740  
Db 5521 ATCCTCACCGATGGCAACGGCAACCGGATCGCGGAGACCAACCGCCCGCGCAGTGGT 5577

Search completed: May 12, 2004, 18:05:33  
Job time : 11480.2 secs



Query Match 40.5%; Score 190.5; DB

A:Molecule type: DNA  
A:Residues: 10-508 <EWL>  
A:Cross-references: NID:G509266; PIDN:CAA82317.1; PID:G509267  
R:Sean, Y.S.; Lee, Y.H.; Pek, U.H.; Kang, H.  
Korean J. Microbiol. 24, 236-242, 1986  
A:Title: Analysis on the nucleotide sequence of the signal region of *Bacillus subtilis* C-101  
A:Reference number: I39803  
A:Accession: I39803  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 10-13, 'V', 15, 'N', 17-18, 'V', 20-21, 'F', 23, 'A', 25-26, 'AI', 29-31, 'P', 33, 'PQ', 36-41, 'PQ', 43-44, 'PQ', 46-47, 'PQ', 49-50, 'PQ', 52-53, 'PQ', 55-56, 'PQ', 58-59, 'PQ', 61-62, 'PQ', 64-65, 'PQ', 67-68, 'PQ', 70-71, 'PQ', 73-74, 'PQ', 76-77, 'PQ', 79-80, 'PQ', 82-83, 'PQ', 85-86, 'PQ', 88-89, 'PQ', 91-92, 'PQ', 94-95, 'PQ', 97-98, 'PQ', 100-101, 'PQ', 103-104, 'PQ', 106-107, 'PQ', 109-110, 'PQ', 112-113, 'PQ', 115-116, 'PQ', 118-119, 'PQ', 121-122, 'PQ', 124-125, 'PQ', 127-128, 'PQ', 130-131, 'PQ', 133-134, 'PQ', 136-137, 'PQ', 139-140, 'PQ', 142-143, 'PQ', 145-146, 'PQ', 148-149, 'PQ', 151-152, 'PQ', 154-155, 'PQ', 157-158, 'PQ', 160-161, 'PQ', 163-164, 'PQ', 166-167, 'PQ', 169-170, 'PQ', 172-173, 'PQ', 175-176, 'PQ', 178-179, 'PQ', 181-182, 'PQ', 184-185, 'PQ', 187-188, 'PQ', 190-191, 'PQ', 193-194, 'PQ', 196-197, 'PQ', 199-200, 'PQ', 202-203, 'PQ', 205-206, 'PQ', 208-209, 'PQ', 211-212, 'PQ', 214-215, 'PQ', 217-218, 'PQ', 220-221, 'PQ', 223-224, 'PQ', 226-227, 'PQ', 229-230, 'PQ', 232-233, 'PQ', 235-236, 'PQ', 238-239, 'PQ', 241-242, 'PQ', 244-245, 'PQ', 247-248, 'PQ', 250-251, 'PQ', 253-254, 'PQ', 256-257, 'PQ', 259-260, 'PQ', 262-263, 'PQ', 265-266, 'PQ', 268-269, 'PQ', 271-272, 'PQ', 274-275, 'PQ', 277-278, 'PQ', 280-281, 'PQ', 283-284, 'PQ', 286-287, 'PQ', 289-290, 'PQ', 292-293, 'PQ', 295-296, 'PQ', 298-299, 'PQ', 301-302, 'PQ', 304-305, 'PQ', 307-308, 'PQ', 310-311, 'PQ', 313-314, 'PQ', 316-317, 'PQ', 319-320, 'PQ', 322-323, 'PQ', 325-326, 'PQ', 328-329, 'PQ', 331-332, 'PQ', 334-335, 'PQ', 337-338, 'PQ', 340-341, 'PQ', 343-344, 'PQ', 346-347, 'PQ', 349-350, 'PQ', 352-353, 'PQ', 355-356, 'PQ', 358-359, 'PQ', 361-362, 'PQ', 364-365, 'PQ', 367-368, 'PQ', 370-371, 'PQ', 373-374, 'PQ', 376-377, 'PQ', 379-380, 'PQ', 382-383, 'PQ', 385-386, 'PQ', 388-389, 'PQ', 391-392, 'PQ', 394-395, 'PQ', 397-398, 'PQ', 400-401, 'PQ', 403-404, 'PQ', 406-407, 'PQ', 409-410, 'PQ', 412-413, 'PQ', 415-416, 'PQ', 418-419, 'PQ', 421-422, 'PQ', 424-425, 'PQ', 427-428, 'PQ', 430-431, 'PQ', 433-434, 'PQ', 436-437, 'PQ', 439-440, 'PQ', 442-443, 'PQ', 445-446, 'PQ', 448-449, 'PQ', 451-452, 'PQ', 454-455, 'PQ', 457-458, 'PQ', 460-461, 'PQ', 463-464, 'PQ', 466-467, 'PQ', 469-470, 'PQ', 472-473, 'PQ', 475-476, 'PQ', 478-479, 'PQ', 481-482, 'PQ', 484-485, 'PQ', 487-488, 'PQ', 490-491, 'PQ', 493-494, 'PQ', 496-497, 'PQ', 499-500, 'PQ', 502-503, 'PQ', 505-506, 'PQ', 508-509, 'PQ', 511-512, 'PQ', 514-515, 'PQ', 517-518, 'PQ', 520-521, 'PQ', 523-524, 'PQ', 526-527, 'PQ', 529-530, 'PQ', 532-533, 'PQ', 535-536, 'PQ', 538-539, 'PQ', 541-542, 'PQ', 544-545, 'PQ', 547-548, 'PQ', 550-551, 'PQ', 553-554, 'PQ', 556-557, 'PQ', 559-560, 'PQ', 562-563, 'PQ', 565-566, 'PQ', 568-569, 'PQ', 571-572, 'PQ', 574-575, 'PQ', 577-578, 'PQ', 580-581, 'PQ', 583-584, 'PQ', 586-587, 'PQ', 589-590, 'PQ', 592-593, 'PQ', 595-596, 'PQ', 598-599, 'PQ', 601-602, 'PQ', 604-605, 'PQ', 607-608, 'PQ', 610-611, 'PQ', 613-614, 'PQ', 616-617, 'PQ', 619-620, 'PQ', 622-623, 'PQ', 625-626, 'PQ', 628-629, 'PQ', 631-632, 'PQ', 634-635, 'PQ', 637-638, 'PQ', 640-641, 'PQ', 643-644, 'PQ', 646-647, 'PQ', 649-650, 'PQ', 652-653, 'PQ', 655-656, 'PQ', 658-659, 'PQ', 661-662, 'PQ', 664-665, 'PQ', 667-668, 'PQ', 670-671, 'PQ', 673-674, 'PQ', 676-677, 'PQ', 679-680, 'PQ', 682-683, 'PQ', 685-686, 'PQ', 688-689, 'PQ', 691-692, 'PQ', 694-695, 'PQ', 697-698, 'PQ', 700-701, 'PQ', 703-704, 'PQ', 706-707, 'PQ', 709-710, 'PQ', 712-713, 'PQ', 715-716, 'PQ', 718-719, 'PQ', 721-722, 'PQ', 724-725, 'PQ', 727-728, 'PQ', 730-731, 'PQ', 733-734, 'PQ', 736-737, 'PQ', 739-740, 'PQ', 742-743, 'PQ', 745-746, 'PQ', 748-749, 'PQ', 751-752, 'PQ', 754-755, 'PQ', 757-758, 'PQ', 760-761, 'PQ', 763-764, 'PQ', 766-767, 'PQ', 769-770, 'PQ', 772-773, 'PQ', 775-776, 'PQ', 778-779, 'PQ', 781-782, 'PQ', 784-785, 'PQ', 787-788, 'PQ', 790-791, 'PQ', 793-794, 'PQ', 796-797, 'PQ', 799-800, 'PQ', 802-803, 'PQ', 805-806, 'PQ', 808-809, 'PQ', 811-812, 'PQ', 814-815, 'PQ', 817-818, 'PQ', 820-821, 'PQ', 823-824, 'PQ', 826-827, 'PQ', 829-830, 'PQ', 832-833, 'PQ', 835-836, 'PQ', 838-839, 'PQ', 841-842, 'PQ', 844-845, 'PQ', 847-848, 'PQ', 850-851, 'PQ', 853-854, 'PQ', 856-857, 'PQ', 859-860, 'PQ', 862-863, 'PQ', 865-866, 'PQ', 868-869, 'PQ', 871-872, 'PQ', 874-875, 'PQ', 877-878, 'PQ', 880-881, 'PQ', 883-884, 'PQ', 886-887, 'PQ', 889-890, 'PQ', 892-893, 'PQ', 895-896, 'PQ', 898-899, 'PQ', 901-902, 'PQ', 904-905, 'PQ', 907-908, 'PQ', 910-911, 'PQ', 913-914, 'PQ', 916-917, 'PQ', 919-920, 'PQ', 922-923, 'PQ', 925-926, 'PQ', 928-

```

|||||: |::: |||||:
Db 419 QMCGNLTHTKFKVTLHKPKQGADTYLE 444

RESULT 5
JN0111
cellulase (EC 3.2.1.4) precursor - Bacillus subtilis (strain BSE616)
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Bacillus subtilis
C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 28-May-1999
C:Accession: JN0111
R:Park, S.H.; Kim, H.K.; Pack, M.Y.
Agric. Biol. Chem. 55, 441-448, 1991
A:Title: Characterization and structure of the cellulase gene of Bacillus subtilis BSE61
A:Reference number: JN0111; MUID:91299280; PMID:1368694
A:Accession: JN0111
A:Molecule type: DNA
A:Residues: 1-499 <PAR>
A:CROSS-references: GB:D01057; NID:g216387; PIDN:BAA00859.1; PID:d1001323; PID:g216388
A:Note: the authors translated the codon ATA for residue 102 as Tyr
C:Comment: The low molecular weight of the mature protein suggests carboxyl-terminal pro
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce
A:Pathway: cellulose degradation
C:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
F:1-29/Domain: signal sequence #status predicted <SIG>

Query Match 40.0%; Score 188; DB 2; Length 499;
Best Local Similarity 40.0%; Pred. No. 6.7e-12;
Matches 34; Conservative 18; Mismatches 31; Indels 2; Gaps 1;

QY 4 GVKVQYKNNDSAPGDNOIKPGLQLVNTGSSVDLSTVTYRVWFTTRDGGSSTLVYNCWAA 63
Db 353 GISVQYRAGDGMNSNQIRPOLQKNNNTVLDKDVARTYRNKNGQNV--DCDYAQ 410

QY 64 MCGGNIRASFGSVNPAFTADTYLQ 88
Db 411 LGCGNVTYKFTLHKPKQGADTYLE 435

RESULT 6
A48954
mannan endo-1,4-beta-mannosidase (EC 3.2.1.78) - Caldocellum saccharolyticum
N:Alternate names: beta-mannanase
C:Species: Caldocellum saccharolyticum
C:Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 08-Oct-1999
R:Gibbs, M.D.; Saul, D.J.; Luthi, E.; Bergquist, P.L.
Appl. Environ. Microbiol. 58, 3864-3867, 1992
A:Title: The beta-mannanase from "Caldocellum saccharolyticum" is part of a multidomain
A:Reference number: A48954; MUID:93119139; PMID:1476429
A:Accession: A48954
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-1331 <GJB>
A:CROSS-references: GB:L01257; NID:g144290; PIDN:AAA71887.1; PID:g144291
A:Note: sequence extracted from NCBI backbone (NCBI:121576, NCBI:121577)
R:Luthi, E.; Bhana Jasmat, N.; Grayling, R.A.; Love, D.R.; Bergquist, P.L.
Appl. Environ. Microbiol. 57, 694-700, 1991
A:Title: Cloning, sequence analysis, and expression in Escherichia coli of a gene coding
A:Reference number: A43745; MUID:91247819; PMID:2039230
A:Accession: B43745
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-337, 'PPROHOHQ' <LUE>
A:CROSS-references: EMBL:M36063; NID:g144292; PIDN:AAA72861.1; PID:g144294
A:Note: the authors translated the codon CAC for residue 262 as Glu
A:Note: this sequence has been revised in reference A48954
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 39.7%; Score 186.5; DB 2; Length 1331;
Best Local Similarity 42.5%; Pred. No. 2.8e-11;
Matches 37; Conservative 18; Mismatches 31; Indels 1; Gaps 1;

```

```

QY 2 SGGVKVQYKNNDSAPGDNOIKPGLQLVNTGSSVDLSTVTYRVWFTTRDGGSSTLVYNCW 61
Db 364 SQGIKVLVYANKETNTTIRPWLKVNSSGSSSIDLSRVTIRYWTYVDGSRQAQSAIS-DW 422

QY 62 AAMGCGNIRASFGSVNPAFTADTYLQ 88
Db 423 AQIGASNVTFKFKLSSSVSGADYYLE 449

RESULT 7
T17120
cellulase (EC 3.2.1.-) precursor, thermoactive - Caldocellum saccharolyticum
C:Species: Caldocellum saccharolyticum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Nov-2000
A:Accession: T17120; A43745
R:Te'o, V.S.; Saul, D.J.; Bergquist, P.L.
Appl. Microbiol. Biotechnol. 43, 291-296, 1995
A:Title: Cella, another gene coding for a multi-domain cellulase from the extreme thermophil
A:Reference number: Z18698; MUID:95336703; PMID:7612247
A:Accession: T17120
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1742 <TEO>
A:CROSS-references: EMBL:L32742; NID:g537499; PID:g537500; PIDN:AAA91086.1
R:Luthi, E.; Bhana Jasmat, N.; Grayling, R.A.; Love, D.R.; Bergquist, P.L.
Appl. Environ. Microbiol. 57, 694-700, 1991
A:Title: Cloning, sequence analysis, and expression in Escherichia coli of a gene coding
A:Reference number: A43745; MUID:91247819; PMID:2039230
A:Accession: A43745
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1516-1544, 'A', 1546-1742 <LUE>
A:CROSS-references: EMBL:M36063; NID:g144292; PIDN:AAA72860.1; PID:g144293
C:Genetics:
A:Gene: cella
C:Keywords: glycosidase; hydrolase

Query Match 39.7%; Score 186.5; DB 2; Length 1742;
Best Local Similarity 42.5%; Pred. No. 3.7e-11;
Matches 37; Conservative 18; Mismatches 31; Indels 1; Gaps 1;

QY 2 SGGVKVQYKNNDSAPGDNOIKPGLQLVNTGSSVDLSTVTYRVWFTTRDGGSSTLVYNCW 61
Db 704 SQGIKVLVYANKETNTTIRPWLKVNSSGSSSIDLSRVTIRYWTYVDGSRQAQSAIS-DW 762

QY 62 AAMGCGNIRASFGSVNPAFTADTYLQ 88
Db 763 AQIGASNVTFKFKLSSSVSGADYYLE 789

RESULT 8
A27198
cellulase (EC 3.2.1.4) precursor - Bacillus subtilis (strain IFO3034)
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Bacillus subtilis
C:Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 15-Oct-1999
C:Accession: A27198
R:Nakamura, A.; Uozumi, T.; Beppu, T.
Eur. J. Biochem. 164, 317-320, 1987
A:Title: Nucleotide sequence of a cellulase gene of Bacillus subtilis.
A:Reference number: A27198; MUID:87190397; PMID:3106035
A:Accession: A27198
A:Molecule type: DNA
A:Residues: 1-499 <NAB>
A:CROSS-references: GB:M28332; NID:g142670; PIDN:AAA22307.1; PID:g142671
A:Experimental source: strain IFO3034
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel
A:Pathway: cellulose degradation
C:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
F:1-36/Domain: signal sequence #status predicted <SIG>

```

Query Match 39.6%; Score 186; DB 2; Length 499;  
Best Local Similarity 40.7%; Pred. No. 1.1e-11;  
Matches 35; Conservative 19; Mismatches 28; Indels 4; Gaps 2;

QY 4 GVKVQYKNDSPAGDNIQKPGQLQVNTGSSVDLSTVTYVWF-TRDGGSSTLVNCDDWA 62  
Db 353 GISVQYKAGDGVNQNIRPQHKNNGNATVDLKDVTARYWYNAKNGQN---FDCDYA 409  
QY 63 AMGCGNIRASFGSVNPNATPTADTYLQ 88  
Db 410 QIGGCVNLTHKFTVHLKPKGADTYLE 435

RESULT 9  
A43802  
cellulase (EC 3.2.1.4) / cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) - Caldocellum  
N:Alternate names: endo-1,4-beta-glucanase  
C:Species: Caldocellum saccharolyticum  
C:Date: 30-Jan-1993 #sequence\_revision 30-Sep-1993 #text\_change 10-Jul-1998  
C:Accession: A43802  
R:Saul, D.J.; Williams, L.C.; Grayling, R.A.; Chamley, L.W.; Love, D.R.; Bergquist, P.L.  
Appl. Environ. Microbiol. 56, 3117-3124, 1990  
A:Title: celB, a gene coding for a bifunctional cellulase from the extreme thermophile  
A:Reference number: A43802; MUID:91136262; PMID:2126700  
A:Accession: A43802  
A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra  
A:Molecule type: DNA  
A:Residues: 1-915 <SAU>  
A:Cross-references: EMBL:X13602  
C:Function:  
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce  
A:Pathway: cellulose degradation  
C:Superfamily: Streptomyces endo-1,4-beta-xylanase A homology  
C:Keywords: glycosidase; hydrolase; polysaccharide degradation  
F:20-320/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SXY>

Query Match 38.8%; Score 182.5; DB 2; Length 915;  
Best Local Similarity 41.9%; Pred. No. 4.8e-11;  
Matches 36; Conservative 18; Mismatches 31; Indels 1; Gaps 1;

QY 3 GVKVQYKNDSPAGDNIQKPGQLQVNTGSSVDLSTVTYVWF-TRDGGSSTLVNCDDWA 62  
Db 349 GQIKVLYANKETNSTTIRPWLKVNSSGSSIDLSRVITRYWYTVGGERAQS- DWA 407  
QY 63 AMGCGNIRASFGSVNPNATPTADTYLQ 88  
Db 408 QIGASNVTFKFKVLSVSGADTYLE 433

RESULT 10  
S02711  
cellulase (EC 3.2.1.4) precursor - Caldocellum saccharolyticum  
N:Alternate names: endo-1,4-beta-glucanase  
C:Contains: cellulase (EC 3.2.1.4); cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91)  
C:Species: Caldocellum saccharolyticum  
C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 08-Oct-1999  
C:Accession: S02711  
R:Saul, D.J.; Williams, L.C.; Love, D.R.; Chamley, L.W.; Bergquist, P.L.  
Nucleic Acids Res. 17, 439, 1989  
A:Title: Nucleotide sequence of a gene from Caldocellum saccharolyticum encoding for exc  
A:Reference number: S02711; MUID:89098398; PMID:2789517  
A:Accession: S02711  
A:Molecule type: DNA  
A:Residues: 1-1039 <SAU>  
A:Cross-references: EMBL:X13602; NID:g40645; PIDN:CAA31936.1; PID:g40646  
C:Genetics:  
A:Gene: celB  
C:Function:  
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce  
A:Pathway: cellulose degradation  
C:Superfamily: Streptomyces endo-1,4-beta-xylanase A homology  
C:Keywords: glycosidase; hydrolase; polysaccharide degradation  
F:1-28/Domain: signal sequence #status predicted <SIG>

F:29-1039/Product: cellulase #status predicted <MAT>  
F:72-373/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SXY>

Query Match 38.8%; Score 182.5; DB 2; Length 1039;  
Best Local Similarity 41.9%; Pred. No. 5.5e-11;  
Matches 36; Conservative 18; Mismatches 31; Indels 1; Gaps 1;

QY 3 GVKVQYKNDSPAGDNIQKPGQLQVNTGSSVDLSTVTYVWF-TRDGGSSTLVNCDDWA 62  
Db 420 GQIKVLYANKETNSTTIRPWLKVNSSGSSIDLSRVITRYWYTVGGERAQS- DWA 478  
QY 63 AMGCGNIRASFGSVNPNATPTADTYLQ 88  
Db 479 QIGASNVTFKFKVLSVSGADTYLE 504

RESULT 11  
A41897  
cellulase homolog - Bacillus lautus (fragment)  
C:Species: Bacillus lautus  
C:Date: 17-Feb-1994 #sequence\_revision 17-Feb-1994 #text\_change 15-Oct-1999  
C:Accession: A41897; S27498  
R:Hansen, C.K.; Didarichsen, B.; Jorgensen, P.L.  
J. Bacteriol. 174, 3522-3531, 1992  
A:Title: celA from Bacillus lautus PL236 encodes a novel cellulose-binding endo-beta-1,4  
A:Reference number: A41897; MUID:92276330; PMID:1592807  
A:Accession: A41897  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-145 <HAN>  
A:Cross-references: EMBL:M76588; NID:g142661; PIDN:AAA22302.1; PID:g142662  
A:Experimental source: PL236  
A:Note: sequence extracted from NCBI backbone (NCBIP:104604)

Query Match 38.4%; Score 180.5; DB 2; Length 145;  
Best Local Similarity 43.9%; Pred. No. 1.1e-11;  
Matches 36; Conservative 18; Mismatches 25; Indels 3; Gaps 2;

QY 7 VOYKNDSPAGDNIQKPGQLQVNTGSSVDLSTVTYVWF-TRDGGSSTLVNCDDWAAMGC 66  
Db 1 LQRAADTNAADNIQKPSFNKNGTSVADLSLKIRYFTKGSAAVNGW-IDWAQLGG 59  
QY 67 GNIRASFGSVNPNATPTADTYLQ 88  
Db 60 SNIQISFG-NHRTGNSDTYVE 79

RESULT 12  
I40548  
bifunctional cellulase precursor - Bacillus sp.  
C:Species: Bacillus sp  
C:Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 15-Oct-1999  
C:Accession: I40548  
R:Han, S.J.; Yoo, Y.J.; Kang, H.S.  
J. Biol. Chem. 270, 26012-26019, 1995  
A:Title: Characterization of a bifunctional cellulase and its structural gene: the cel  
A:Reference number: I40548; MUID:96029707; PMID:7532793  
A:Accession: I40548  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-486 <RES>  
A:Cross-references: EMBL:U27084; NID:g857575; PIDN:AAC43478.1; PID:g857576  
C:Genetics:  
A:Gene: cel

Query Match 38.3%; Score 180; DB 2; Length 486;  
Best Local Similarity 38.8%; Pred. No. 4.5e-11;  
Matches 33; Conservative 18; Mismatches 32; Indels 2; Gaps 1;

QY 4 GVKVQYKNDSPAGDNIQKPGQLQVNTGSSVDLSTVTYVWF-TRDGGSSTLVNCDDWA 63  
Db 353 GISVQYKAGDGVNQNIRPQHKNNGNATVDLKDVTARYWYNAKNGQN---DCCDYEQ 410

```
Qy 64 MCGNIRASFGSVNPATPTADTYLQ 88
Db 411 LGCGNVSHVTTLHKPKQGADTYLE 435

RESULT 13
B41897
cellulase (EC 3.2.1.4) - Bacillus lautus
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Bacillus lautus
C:Date: 10-Mar-1994 #sequence_revision 10-Mar-1994 #text_change 15-Oct-1999
C:Accession: B41897; S27499
R:Hansen, C.K.; Diderichsen, B.; Jorgensen, P.L.
J. Bacteriol. 174, 3522-3531, 1992
A:Title: celsA from Bacillus lautus PL236 encodes a novel cellulose-binding endo-beta-1,4
A:Reference number: A41897; MUID:92276330; PMID:1592807
A:Accession: B41897
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-700 <HAN>
A:Cross-references: EMBL:M76588; NID:g142661; PIDN:AAA22303.1; PID:g142663
A:Experimental source: PL236
A:Note: sequence extracted from NCBI backbone (NCBIP:104605)
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce
A:Pathway: cellulose degradation
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 36.6%; Score 172; DB 2; Length 700;
Best Local Similarity 44.9%; Pred. No. 4.5e-10;
Matches 40; Conservative 18; Mismatches 27; Indels 4; Gaps 3;

Qy 1 VSGGVKVOYKND-SAPGDNOIKPGLQVNTGSSVDLSTVTYVWFTRDGGSSLTLYNC 59
Db 549 VNSDLVQYKDDNNATDNOIKPHFNQNGTSFVDSLTLRYFTKD-SSAAMNGWI 607

Qy 60 DWAMCGCNIRASFGSVNPATPTADTYLQ 88
Db 608 DWAKLGGSNQISFGNHGA--DSDTYAE 634

RESULT 14
A47704
endoglucanase I (EC 3.2.1.-) Cell - Clostridium thermocellum
C:Species: Clostridium thermocellum
C:Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 15-Oct-1999
C:Accession: A47704
R:Hazlewood, G.P.; Davidson, K.; Laurie, J.I.; Huskisson, N.S.; Gilbert, H.J.
J. Gen. Microbiol. 139, 307-316, 1993
A:Title: Gene sequence and properties of CellI, a family E endoglucanase from Clostridium
A:Reference number: A47704; MUID:93171873; PMID:8436949
A:Accession: A47704
A>Status: preliminary
A:Molecule type: DNA; protein
A:Residues: 1-879 <HAZ>
A:Cross-references: GB:L04735; NID:g144807; PIDN:AAA20892.1; PID:g144808
A:Note: sequence extracted from NCBI backbone (NCBIN:125637, NCBIP:125638)
C:Keywords: glycosidase; hydrolase

Query Match 35.2%; Score 165.5; DB 2; Length 879;
Best Local Similarity 33.0%; Pred. No. 2.7e-09;
Matches 29; Conservative 26; Mismatches 32; Indels 1; Gaps 1;

Qy 1 VSGGVKVOYKND-SAPGDNOIKPGLQVNTGSSVDLSTVTYVWFTRDGGSSLTLYNC 60
Db 736 IKGEVTLQYANGAGATSNINPRPKIINNGTKAINLSDVKIRYYTKEGGASQNFW-CD 794

Qy 61 WAAMCGCNIRASFGSVNPATPTADTYLQ 88
Db 795 WSSAGNGNVGTGNFFNLSPKEGADTCLE 822

RESULT 15
```

```
S36859
cipA protein - Clostridium thermocellum
N:Alternate names: probable cellulosome protein large chain SL
C:Species: Clostridium thermocellum
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 01-Dec-2000
C:Accession: S36859; S33527; S25767; S28659; T18261
R:Gerngross, U.T.; Demain, A.L.
submitted to the EMBL Data Library, January 1993
A:Reference number: S36859
A:Accession: S36859
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1854 <GER>
A:Cross-references: EMBL:L08665
R:Gerngross, U.T.; Romaniec, M.P.M.; Kobayashi, T.; Huskisson, N.S.; Demain, A.L.
Mol. Microbiol. 8, 325-334, 1993
A:Title: Sequencing of a Clostridium thermocellum gene (cipA) encoding the cellulosomal
A:Reference number: S33527; MUID:93302508; PMID:8316083
A:Accession: S33527
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1615,1617-1854 <GE2>
A:Cross-references: EMBL:L08665
R:Fujino, T.; Beguin, P.; Aubert, J.P.
FEMS Microbiol. Lett. 94, 165-170, 1992
A:Title: Cloning of a Clostridium thermocellum DNA fragment encoding polypeptides that b
A:Reference number: S25767
A:Accession: S25767
A:Molecule type: DNA
A:Residues: 1216-1232, 'Y', 1234-1241, 'T', 1243-1319, 'R', 1321-1615,1617-1854 <FUJ>
A:Cross-references: EMBL:X67406
R:Bequin, P.
submitted to the EMBL Data Library, August 1992
A:Reference number: S28659
A:Accession: S28659
A:Molecule type: DNA
A:Residues: 1216-1232, 'Y', 1234-1241, 'T', 1243-1319, 'R', 1321-1615,1617-1768, 'R', 1770-1854
A:Cross-references: EMBL:X67406
R:Fujino, T.; Bequin, P.; Aubert, J.P.
J. Bacteriol. 175, 1891-1899, 1993
A:Title: Organization of a Clostridium thermocellum gene cluster encoding the cellulosome
e.
A:Reference number: Z18847; MUID:93209931; PMID:8458832
A:Accession: T18261
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1821-1854 <FU2>
A:Cross-references: EMBL:X67506; NID:g296879; PID:g296880; PIDN:CAA47840.1
C:Genetics:
A:Gene: cipa

Query Match 35.1%; Score 165; DB 2; Length 1854;
Best Local Similarity 36.1%; Pred. No. 6.9e-09;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

Qy 1 VSGGVKVOYKND-SAPGDNOIKPGLQVNTGSSVDLSTVTYVWFTRDGGSSLTLYNC 60
Db 365 VSGNLKVFYNSNPSTTSINPQFKVTNTGSSAIDLSKLTLYRYTVDGQKQDTFW-CD 423

Qy 61 WAAM-----GCGNIRASFGSVNPATPTADTYLQ 88
Db 424 HAAITGNSGNYGITSNVKGTFFVKMSSSTNNADTYLE 460

Search completed: May 11, 2004, 12:10:26
Job time : 6.17339 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 11:41:35 ; Search time 2.42639 seconds  
(without alignments)  
1909.933 Million cell updates/sec

Title: US-09-917-376-4

Perfect score: 470

Sequence: 1 VSGGVKQYKNDSPAGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	190	40.4	499	1 GUN1_BACSU	P07983 bacillus su
2	190	40.4	499	1 GUN2_BACSU	P10475 bacillus su
3	188	40.0	499	1 GUN3_BACSU	P23549 bacillus su
4	186.5	39.7	1331	1 MANB_CALSA	P22533 caldocellum
5	186.5	39.7	1742	1 GUNA_CALSA	P22534 caldocellum
6	182.5	38.8	1039	1 GUNB_CALSA	P10474 c endogluc
7	180.5	38.4	145	1 YCEA_PAEIA	P29718 paenibacill
8	172	36.6	700	1 GUNA_PAEIA	Q02934 clostridium
9	165.5	35.2	879	1 GUN1_CLOTM	Q01866 clostridium
10	165	35.1	772	1 CIPB_CLOTM	Q06851 clostridium
11	165	35.1	1853	1 CIPA_CLOTM	Q59394 erwinia car
12	149.5	31.8	444	1 GUNW_ERWCA	Q47096 erwinia car
13	149.5	31.8	505	1 GUNW_ERWCA	Q59395 erwinia car
14	147.5	31.4	504	1 GUNW_ERWCA	P50900 clostridium
15	147.5	31.4	914	1 GUX2_CLOS	P23659 clostridium
16	141.5	30.1	986	1 GUNZ_CLOS	P38058 clostridium
17	114	24.3	1848	1 CBPA_CLOCL	P01809 mus musculu
18	67.5	14.4	118	1 HV19_MOUSE	P75316 methanococc
19	67.5	14.4	215	1 FLAI_METVA	Q9fsg7 malus domes
20	66.5	14.1	246	1 TPIA_MALDO	P71553 m bifunctio
21	65	13.8	523	1 PUR9_MYCTU	Q8rdj9 thermoanaer
22	64.5	13.7	556	1 ILVD_THETN	P32660 saccharomyc
23	64	13.6	1571	1 ATCS_YEAST	P57719 thermoplasma
24	63.5	13.5	247	1 FLAI_THEVO	O17446 schistosoma
25	63.5	13.5	465	1 TX3H_SCHNA	Q9upv9 homo sapien
26	63	13.4	953	1 O106_HUMAN	Q15648 h peroxisom
27	63	13.4	1581	1 PRXB_HUMAN	P33693 rhizobium m
28	62.5	13.3	269	1 EXOK_RHIME	P29352 mus musculu
29	62.5	13.3	802	1 PTN8_MOUSE	Q99n50 mus musculu
30	62	13.2	950	1 STL2_MOUSE	P09815 pseudomonas
31	62	13.2	1210	1 ICEN_PSEFL	P50937 rhodobacter
32	61.5	13.1	252	1 HIS6_RHOSH	P38108 pseudomonas
33	61.5	13.1	316	1 MUCB_PSEAE	

34	61.5	13.1	328	1 HAIQ_MOUSE	P14428 mus musculu
35	61.5	13.1	368	1 HAIW_MOUSE	P03991 mus musculu
36	61.5	13.1	489	1 GATA_CAUCR	Q9a510 caulobacter
37	61.5	13.1	678	1 YIHQ_ECOLI	P32138 escherichia
38	61	13.0	120	1 HV5Q_MOUSE	P06329 mus musculu
39	61	13.0	338	1 RTCA_ECOLI6	Q8fcs8 escherichia
40	61	13.0	338	1 RTCA_ECOLI	P46849 escherichia
41	61	13.0	342	1 RTCA_ECO57	P58127 escherichia
42	60.5	12.9	1157	1 C9CA_BACTO	Q45733 bacillus th
43	60	12.8	312	1 LECF_ALEAU	P18891 aleuria aur
44	60	12.8	1226	1 PAT2_CAEEL	P34446 caenorhabdi
45	59.5	12.7	261	1 COTN_BACSU	P54507 bacillus su

#### ALIGNMENTS

RESULT 1  
GUN1\_BACSU STANDARD; PRT; 499 AA.  
ID GUN1\_BACSU  
AC P07983;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-NOV-1991 (Rel. 20, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)  
DE (Cellulase).  
GN BGLC OR GLD.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DLG;  
RX MEDLINE=87194581; PubMed=3106328;  
RA Robson L.M.; Chambliss G.H.;  
RT "Endo-beta-1,4-glucanase gene of Bacillus subtilis DLG.";  
RL J. Bacteriol. 169:2017-2025(1987).  
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic linkages in cellulose, lichenin and cereal beta-D-glucans.  
CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl hydrolases).

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; M16185; AAA22496.1; ALT\_INIT.  
FIR; A26874; A26874.  
HSSP; O85465; IA3H.  
InterPro; IPR001956; CBD 3.  
InterPro; IPR008965; Cellul\_bind.  
InterPro; IPR001547; Glyco\_Hydro\_5.  
Pfam; PF00942; CBM\_3; 1.  
Pfam; PF00150; cellulase; 1.  
ProDom; PD001947; CBD 3; 1.  
PROSITE; PS00659; GLYCOSYL HYDROL F5; 1.  
Cellulose degradation; Hydrolase; Glycosidase; Signal.  
FT SIGNAL 1 29  
FT CHAIN 30 499 ENDOGLUCANASE.  
FT ACT SITE 169 169 PROTON DONOR (BY SIMILARITY).  
FT ACT SITE 257 257 NUCLEOPHILE (BY SIMILARITY).  
FT DOMAIN 350 499 CELLULOSE-BINDING (BY SIMILARITY).  
SQ SEQUENCE 499 AA; 55187 MW; 339D04EB95A63BE1 CRC64;

Query Match 40.4%; Score 190; DB 1; Length 499;  
Best Local Similarity 43.0%; Pred. No. 1.3e-13;  
Matches 37; Conservative 17; Mismatches 28; Indels 4; Gaps 2;

QY 4 GVKVQYKNDSPAGDNQIKPGLQVNTGSSVDLSITVTYWF-TRDGGSTLYVNCWA 62

```

Db 353 GVSQYKAGDGRVNSQIRPOLKRNKGATVDLKDVTARYWYKKNQGN---FDCDYA 409
QY 63 AMCGNIRASFGSVNPTADTYLQ 88
Db 410 QMGCGNLTHKFTVTLHKPKQGAITYLE 435

RESULT 2
GUN2_BACSU
ID GUN2_BACSU STANDARD; PRT; 499 AA.
AC P10475;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Carboxymethyl-cellulase) (CMCASE) (Cellulase).
GN BGLC OR GLD OR EGLS OR BSU18130.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PAP115;
RX MEDLINE=87066783; PubMed=30241130;
RA Mackay R.M., Lo A., Willick G., Zuker M., Baird S., Dove M.,
RA Moraneli F., Seligy V.;
RT "Structure of a Bacillus subtilis endo-beta-1,4-glucanase gene.";
RL Nucleic Acids Res. 14:9159-9170(1986).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CK-2;
RX MEDLINE=95225656; PubMed=7710280;
RA Lindahl V., Aa K., Tronsmo A.;
RT "Nucleotide sequence of an endo-beta-1,4-glucanase gene from Bacillus
RT subtilis CK-2.";
RL Antonie Van Leeuwenhoek 66:327-332(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=97124194; PubMed=8969507;
RA Rose M., Entian K.D.;
RT "New genes in the 170 degrees region of the Bacillus subtilis genome
RT encode DNA gyrase subunits, a thioredoxin, a xylanase and an amino
RT acid transporter.";
RL Microbiology 142:3097-3101(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Boursier L., Bessieres P., Bolotin A., Borchert S.,
RA Borriess R., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerston I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Frits C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Chim S.Y., Glaser P., Goffeau A., Colightly E.J., Grandi G.,
RA Guiseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Kjaer-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic C., Ruchelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,

```

```

RA Takeuchi M., Tanakoshi A., Tanaka T., Terpetra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumatein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256(1997).
RN [5]
RP SEQUENCE OF 30-45.
RC STRAIN=CK-2;
RX MEDLINE=95225655; PubMed=7710279;
RA Aa K., Flengsrud R., Lindahl V., Tronsmo A.;
RT "Characterization of production and enzyme properties of an
RT endo-beta-1,4-glucanase from Bacillus subtilis CK-2 isolated from
RT compost soil.";
RL Antonie Van Leeuwenhoek 66:319-326(1994).
CC -| CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -| SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl
CC hydrolases).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z29076; CAA82317.1; -;
DR EMBL; X04689; CAA28392.1; -;
DR EMBL; X67044; CAA47429.1; -;
DR EMBL; Z73234; CAA97610.1; ALT INIT.
DR EMBL; Z99113; CAB13696.1; ALT_INIT.
DR PIR; G69593; G69593.
DR HSSP; O85465; 1A3H.
DR Subtilist; BG10437; bg1C.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR001547; Glyco_Hydro_5.
DR Pfam; PF00942; CBM_3; 1.
DR Pfam; PF00150; cellulase; 1.
DR ProDom; PD001947; CBD_3; 1.
DR PROSITE; PS00659; GLYCOSYL HYDROL F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal;
KW Complete proteome.
FT SIGNAL 1 29
FT CHAIN 30 499
FT ACT_SITE 169 169
FT ACT_SITE 257 257
FT DOMAIN 350 499
FT CONFLICT 283 283
SQ SEQUENCE 499 AA; 55287 MW; 8F735FE711B3AE2 CRC64;

Query Match 40.4%; Score 190; DB 1; Length 499;
Best Local Similarity 40.7%; Pred. No. 13e-13;
Matches 35; Conservative 21; Mismatches 26; Indels 4; Gaps 2;

QY 4 GVKVQYKNNDSAPGDNQIRPOLKRNKGATVDLKDVTARYWYKKNQGN---FDCDYA 62
Db 353 GVSQYKAGDGRVNSQIRPOLKRNKGATVDLKDVTARYWYKKNQGN---FDCDYA 409
QY 63 AMCGNIRASFGSVNPTADTYLQ 88
Db 410 QMGCGNLTHKFTVTLHKPKQGAITYLE 435

RESULT 3
GUN3_BACSU
ID GUN3_BACSU STANDARD; PRT; 499 AA.
AC P23549;
DT 01-NOV-1991 (Rel. 20, Created)

```

DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE 1 Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-giucanase)  
 DE (Carboxymethyl-cellulase) (CMCSE) (Cellulase).  
 GN BGLC.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BS616;  
 RX MEDLINE=91299280; PubMed=1368694;  
 RA Park S.H., Kim H.K., Park M.Y.;  
 RT "Characterization and structure of the cellulase gene of Bacillus  
 RT subtilis BSE616";  
 RL Agric. Biol. Chem. 55:441-448(1991).  
 CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
 CC linkages in cellulose, lichenin and cereal beta-D-glucans.  
 CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl  
 CC hydrolases).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; D01057; BAA00859.1; -;  
 DR PIR; JN0111; JN0111.  
 DR HSP; O85465; 1A3H.  
 DR InterPro; IPR001956; CBD\_3.  
 DR InterPro; IPR008965; Cellul\_bind.  
 DR InterPro; IPR001547; Glyco\_hydro\_5.  
 DR Pfam; PF00942; CBM\_3; 1.  
 DR Pfam; PF00150; cellulase; 1.  
 DR ProDom; PD001947; CBD\_3; 1.  
 DR PROSITE; PS00659; GLYCOSYL\_HYDROL\_F5; 1.  
 DR Cellulose degradation; Hydrolase; Glycosidase; Signal.  
 KW SIGNAL 1 29  
 FT CHAIN 30 499 ENDOGLUCANASE.  
 FT ACT\_SITE 169 169 PROTON DONOR (BY SIMILARITY).  
 FT ACT\_SITE 257 257 NUCLEOPHILE (BY SIMILARITY).  
 FT DOMAIN 350 499 CELLULOSE-BINDING (BY SIMILARITY).  
 SQ SEQUENCE 499 AA; 55169 MW; 2E821E3DB8BACA04 CRC64;  
 Query Match 40.0%; Score 188; DB 1; Length 499;  
 Best Local Similarity 40.0%; Pred. No. 2.2e-13;  
 Matches 34; Conservative 18; Mismatches 31; Indels 2; Gaps 1;  
 OY 4 GVKVQYKXNDSPGDNQIKPGLQVNTGSSVDLSTVTYRYWFTDGGSSTLVYNCDMAA 63  
 Db :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:  
 353 GISVQYRAGSGMNSGNIRPQIKNGNTTVDLKDVTARYWYNAKKNQNV--DCDYAQ 410  
 OY 64 MCGNIRASFGSVNPTPTADTYLQ 88  
 Db :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:  
 411 LGCGNVYKFTVTLHKPGKQADTYLE 435  
 RESULT 4  
 MANB\_CALSA STANDARD; PRT; 1331 AA.  
 AC P22533;  
 DT 01-AUG-1991 (Rel. 19, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Beta-mannanase/endoglucanase A precursor [Includes: Mannan endo-1,4-  
 DE beta-mannosidase A (EC 3.2.1.78) (Beta-mannanase) (Endo-1,4-  
 DE mananase); Endo-1,4-beta-giucanase (EC 3.2.1.4) (Cellulase)].  
 GN MANA.  
 OS Caldocellum saccharolyticum (Caldicellulosiruptor saccharolyticus).  
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;

OC Caldicellulosiruptor.  
 OX NCBI\_TaxID=44001;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93119139; PubMed=1476429;  
 RA Gibbs M.D., Saul D.J., Luthi E., Bergquist P.L.;  
 RT "The beta-mannanase from 'Caldocellum saccharolyticum' is part of a  
 RT multidomain enzyme";  
 RL Appl. Environ. Microbiol. 58:3864-3867(1992).  
 RN [2]  
 RP SEQUENCE OF 1-346 FROM N.A.  
 RX MEDLINE=91247819; PubMed=2039230;  
 RA Luechi E., Jasmat N.B., Grayling R.A., Love D.R., Bergquist P.L.;  
 RT "Cloning, sequence analysis, and expression in Escherichia coli of a  
 RT gene coding for a beta-mannanase from the extremely thermophilic  
 RT bacterium 'Caldocellum saccharolyticum'";  
 RL Appl. Environ. Microbiol. 57:694-700(1991).  
 CC -!- FUNCTION: DEGRADATION OF HEMICELLULOSES, THE SECOND MOST ABUNDANT  
 CC POLYSACCHARIDES IN NATURE. CONTAINS TWO CATALYTIC DOMAINS WITH  
 CC MANNANASE AND ENDOGLUCANASE ACTIVITIES.  
 CC -!- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-beta-D-mannosidic  
 CC linkages in mannans, galactomannans, glucomannans, and  
 CC galactoglucomannans.  
 CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
 CC linkages in cellulose, lichenin and cereal beta-D-glucans.  
 CC -!- MISCELLANEOUS: This enzyme is most active at pH 6 and 80 degrees  
 CC Celsius.  
 CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY  
 CC A (FAMILY 5 OF GLYCOSYL HYDROLASES).  
 CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY  
 CC J (FAMILY 44 OF GLYCOSYL HYDROLASES).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; L01257; AAA71887.1; -;  
 DR EMBL; M36063; AAA72861.1; -;  
 DR PIR; A48954; A48954.  
 DR HSP; O06851; INBC.  
 DR InterPro; IPR001956; CBD\_3.  
 DR InterPro; IPR008965; Cellul\_bind.  
 DR InterPro; IPR001547; Glyco\_hydro\_5.  
 DR Pfam; PF00942; CBM\_3; 2.  
 DR Pfam; PF00150; cellulase; 1.  
 DR ProDom; PD001947; CBD\_3; 2.  
 DR PROSITE; PS00659; GLYCOSYL\_HYDROL\_F5; 1.  
 DR Hydrolase; Glycosidase; Cellulose degradation; Signal;  
 KW Multifunctional enzyme.  
 FT SIGNAL 1 41  
 FT CHAIN 42 1331 BETA-MANNANASE/ENDOGLUCANASE A.  
 FT DOMAIN 42 325 CATALYTIC (MANNANASE ACTIVITY).  
 FT DOMAIN 326 361 PRO/SER/THR-RICH (PT BOX).  
 FT DOMAIN 362 518 SUBSTRATE-BINDING (POTENTIAL).  
 FT DOMAIN 519 564 PRO/SER/THR-RICH (PT BOX).  
 FT DOMAIN 565 720 SUBSTRATE-BINDING (POTENTIAL).  
 FT DOMAIN 721 780 PRO/SER/THR-RICH (PT BOX).  
 FT DOMAIN 781 1331 CATALYTIC (ENDOGLUCANASE ACTIVITY).  
 FT ACT\_SITE 162 162 PROTON DONOR (BY SIMILARITY).  
 FT ACT\_SITE 257 257 NUCLEOPHILE (BY SIMILARITY).  
 FT CONFLICT 338 338 T -> P (IN REF. 2).  
 FT CONFLICT 340 346 TPPTPT -> ROHQHQ (IN REF. 2).  
 SQ SEQUENCE 1331 AA; 146892 MW; PFBCA51BB8DBF0E0 CRC64;  
 Query Match 39.7%; Score 186.5; DB 1; Length 1331;  
 Best Local Similarity 42.5%; Pred. No. 1e-12;  
 Matches 37; Conservative 18; Mismatches 31; Indels 1; Gaps 1;  
 OY 2 SGGVKYQXNDSPGDNQIKPGLQVNTGSSVDLSTVTYRYWFTDGGSSTLVYNCDW 61



```
Db 364 SQGKVLKANKETNTTIRPWLKVNSSGSSIDLSRVITRYWYTVGDSSTLVNCDW 422
QY 62 AAMGCGNIRASFGSVNPAFTPTADTYLQ 88
Db 423 AQIGASNVTFKFKLSSSVSGADYLYE 449

RESULT 5
GUNA_CALSA
ID GUNA_CALSA STANDARD; PRT; 1742 AA.
AC P22534;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase A)
DE (Cellulase A).
GN CELA.
OS Caldocellum saccharolyticum (Caldicellulosiruptor saccharolyticus).
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=44001;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95336703; PubMed=7612247;
RA Te'O V.S., Saul D.J., Bergquist P.L.;
RT "celA, another gene coding for a multidomain cellulase from the
RT extreme thermophile Caldocellum saccharolyticum.";
RL Appl. Microbiol. Biotechnol. 43:291-296(1995).
RN [2]
RP SEQUENCE OF 1516-1742 FROM N.A.
RX MEDLINE=91247819; PubMed=2039230;
RA Leuthi E., Jasmat N.B., Grayling R.A., Love D.R., Bergquist P.L.;
RT "Cloning, sequence analysis, and expression in Escherichia coli of a
RT gene coding for a beta-mannanase from the extremely thermophilic
RT bacterium 'Caldocellum saccharolyticum'.";
RL Appl. Environ. Microbiol. 57:694-700(1991).
CC -!- FUNCTION: THE N-TERMINAL DOMAIN OF CELA ENCODES FOR AN
CC ENDOGLUCANASE ACTIVITY ON CARBOXYMETHYLCELLULOSE. THE C-TERMINAL
CC DOMAIN PROBABLY ACT SYNERGISTICALLY TO HYDROLYZE CRYSTALLINE
CC CELLULOSE.
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- PTM: The linker region (also termed "hinge") may be a potential
CC site for proteolysis.
CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
CC E (FAMILY 9 OF GLYCOSYL HYDROLASES).
CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
CC L (FAMILY 48 OF GLYCOSYL HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L32742; AAA91086.1; -
CC EMBL; M36063; AAA72860.1; -
CC EMBL; L01257; -; NOT_ANNOTATED_CDS.
CC PIR; T17120; T17120.
CC HSSP; P26221; 1TF4.
CC InterPro; IPR001956; CBD_3.
CC InterPro; IPR008965; Cellul_bind.
CC InterPro; IPR000556; Glyco_hydro_48.
CC InterPro; IPR001701; Glyco_hydro_9.
CC InterPro; IPR008928; Glyco_trans_6hp.
CC Pfam; PF00942; CBM_3; 3.
CC Pfam; PF02011; Glyco_hydro_48; 1.
CC Pfam; PF00759; Glyco_hydro_9; 1.
CC PRINTS; PR00844; GLHYDRLASE48.
CC ProDom; PD001947; CBD_3; 2.
```

```
DR ProDom; PD011903; Glyco_hydro_48; 1.
DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal; Repeat.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 1742 ENDOGLUCANASE A.
FT DOMAIN 24 642 CATALYTIC 1.
FT DOMAIN 643 700 LINKER ("HINGE") (PRO-THR BOX).
FT DOMAIN 701 857 CELLULOSE-BINDING (BY SIMILARITY).
FT DOMAIN 858 903 LINKER ("HINGE") (PRO-THR BOX).
FT DOMAIN 904 1060 CELLULOSE-BINDING (BY SIMILARITY).
FT DOMAIN 1061 1112 LINKER ("HINGE") (PRO-THR BOX).
FT DOMAIN 1113 1742 CATALYTIC 2.
FT ACT_SITE 396 396 BY SIMILARITY.
FT ACT_SITE 434 434 BY SIMILARITY.
FT ACT_SITE 443 443 BY SIMILARITY.
FT CONFLICT 1545 1545 T -> A (IN REF. 2).
SQ SEQUENCE 1742 AA; 193696 MW; 3F0699A2123EED07 CRC64;

Query Match 39.7%; Score 186.5; DB 1; Length 1742;
Best Local Similarity 42.5%; Pred. No. 1.4e-12;
Matches 37; Conservative 18; Mismatches 31; Indels 1; Gaps 1;

QY 2 SGGVKQYKNDSPAGDGNQIKPGLQLVNTGSSVDLSVTIVRYWFTRDGSGSTLVNCDW 61
Db 704 SQGKVLKANKETNTTIRPWLKVNSSGSSIDLSRVITRYWYTVGDSSTLVNCDW 762
QY 62 AAMGCGNIRASFGSVNPAFTPTADTYLQ 88
Db 763 AQIGASNVTFKFKLSSSVSGADYLYE 789

RESULT 6
GUNA_CALSA
ID GUNA_CALSA STANDARD; PRT; 1039 AA.
AC P10474;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Endoglucanase/exoglucanase B precursor [Includes: Endoglucanase
DE (EC 3.2.1.4) (Endo-1,4-beta-glucanase) (Cellulase)
DE (Cellulohydrolase); Exoglucanase (EC 3.2.1.91) (Exocellulohydrolase)
DE (1,4-beta-cellobiohydrolase)].
GN CELB.
OS Caldocellum saccharolyticum (Caldicellulosiruptor saccharolyticus).
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=44001;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89098398; PubMed=2789517;
RA Saul D.J., Williams L.C., Love D.R., Chamley I.W., Bergquist P.L.;
RT "Nucleotide sequence of a gene from Caldocellum saccharolyticum
RT encoding for exocellulase and endocellulase activity.";
RL Nucleic Acids Res. 17:439-439(1989).
CC -!- FUNCTION: THIS PROTEIN IS MADE UP OF TWO DOMAINS: THE N-TERMINAL
CC DOMAIN HAS EXOGLUCANASE ACTIVITY WHILE THE C-TERMINAL DOMAIN IS
CC AN ENDOGLUCANASE.
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC in cellulose and cellotetraose, releasing cellobiose from the non-
CC reducing ends of the chains.
CC -!- SIMILARITY: In the N-terminal section; belongs to cellulase family
CC F (family 10 of glycosyl hydrolases).
CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
CC A (FAMILY 5 OF GLYCOSYL HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC modified and this statement is not removed. Usage by and for commercial
```

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
-----  
DR EMBL; X13602; CRA31936.1; .  
DR PIR; S02711; S02711.  
DR HSSP; Q06851; INBC.  
DR InterPro; IPR001956; CBD 3.  
DR InterPro; IPR008965; Cellul\_bind.  
DR InterPro; IPR001000; Glyco\_hydro\_10.  
DR InterPro; IPR001547; Glyco\_hydro\_5.  
DR Pfam; PF00942; CEM 3; 1.  
DR Pfam; PF00150; cellulase; 1.  
DR Pfam; PF00331; Glyco\_hydro\_10; 1.  
DR PRINTS; PR00134; GLHYDRLASE10.  
DR PRODOM; PD001947; CBD 3; 1.  
DR SMART; SM00633; Glyco\_10; 1.  
DR PROSITE; PS00591; GLYCOSYL\_HYDROL\_F10; 1.  
DR PROSITE; PS00659; GLYCOSYL\_HYDROL\_F5; 1.  
DR Cellulose degradation; Hydrolase; Glycosidase; Repeat;  
KW Multifunctional enzyme; Signal.  
FT SIGNAL 1 28  
FT CHAIN 29 1039 ENDOGLUCANASE/EXOGLUCANASE B.  
FT DOMAIN 376 416 THR/PRO-RICH, TANDEM REPEATS OF T-P.  
FT DOMAIN 417 570 CELLULOSE-BINDING (BY SIMILARITY).  
FT DOMAIN 571 618 THR/PRO-RICH, TANDEM REPEATS OF T-P.  
FT ACT SITE 177 177 PROTON DONOR (POTENTIAL).  
FT ACT SITE 285 285 NUCLEOPHILE (BY SIMILARITY).  
FT ACT SITE 792 792 BY SIMILARITY.  
SQ SEQUENCE 1039 AA; 117641 MW; 0E0378171594DDAE CRC64;  
  
Query Match 38.8%; Score 182.5; DB 1; Length 1039;  
Best Local Similarity 41.9%; Pred. No. 2.1e-12;  
Matches 36; Conservative 18; Mismatches 31; Indels 1; Gaps 1;  
  
QY 3 GGKVKQYKNDSPAGDNIKPGQLVNTGSSVDLSVTYVYFTRDGGSTLVNCDWA 62  
DB 420 GQIKVLYANKETNTTIRPWLKVNSSGSSIDLSRTYRTYVTDGERAQSAVS-DWA 478  
  
QY 63 AMGCVNIRASFGSNVPATPTADTYLQ 88  
DB 479 QIGASNVTFKFKLSSVSGADYLYLE 504  
  
RESULT 7  
YCEA\_PAEIA STANDARD; PRT; 145 AA.  
AC Y29718;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein in celsA 5' region (Fragment).  
OS Paenibacillus lautus (Bacillus lautus).  
OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.  
OX NCBI\_TaxID=1401;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PL236;  
RX MEDLINE=92276330; PubMed=1592807;  
RA Hansen C.K., Joergensen P.L., Diderichsen B.;  
RT "celsA from Bacillus lautus PL236 encodes a novel cellulose-binding  
endo-beta-1,4-glucanase."  
RL J. Bacteriol. 174:3522-3531(1992).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
-----  
DR EMBL; M76588; AAA22302.1; .  
DR PIR; A41897; A41897.  
DR HSSP; Q06851; INBC.  
DR InterPro; IPR001956; CBD 3.  
DR Pfam; PF00942; CEM 3; 1.  
DR PRODOM; PD001947; CBD 3; 1.  
DR Cellulose degradation; Hydrolase; Glycosidase; Signal.  
FT SIGNAL 1 33 POTENTIAL.  
FT CHAIN 34 700 ENDOGLUCANASE A.  
FT DOMAIN 548 700 CELLULOSE-BINDING (BY SIMILARITY).  
FT ACT SITE 213 213 BY SIMILARITY.  
SQ SEQUENCE 700 AA; 76910 MW; 3D5C8CADAS3EEB0F CRC64;  
  
Query Match 36.6%; Score 172; DB 1; Length 700;  
Best Local Similarity 43.9%; Pred. No. 3.4e-13;  
Matches 36; Conservative 18; Mismatches 25; Indels 3; Gaps 2;  
  
QY 7 VOYKNDSPAGDNIKPGQLVNTGSSVDLSVTYVYFTRDGGSTLVNCDWAAMGC 66  
DB 1 LQYRAADTNAADNQIKPSFNKNNGTSAVDLSLKIRYFTKDGSAVNGW-IDWAQLGG 59  
  
QY 67 GNIRASFGSNVPATPTADTYLQ 88  
DB 60 SNIQISFG--NHTGTNSDTYVE 79  
  
RESULT 8  
GUNA\_PAEIA STANDARD; PRT; 700 AA.  
ID GUNA\_PAEIA  
AC P29719;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)  
DE (Cellulase A) (EG-A).  
GN CELA.  
OS Paenibacillus lautus (Bacillus lautus).  
OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.  
OX NCBI\_TaxID=1401;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PL236;  
RX MEDLINE=92276330; PubMed=1592807;  
RA Hansen C.K., Diderichsen B., Joergensen P.L.;  
RT "celsA from Bacillus lautus PL236 encodes a novel cellulose-binding  
endo-beta-1,4-glucanase."  
RL J. Bacteriol. 174:3522-3531(1992).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
-----  
DR EMBL; M76588; AAA22303.1; .  
DR PIR; B41897; B41897.  
DR HSSP; Q06851; INBC.  
DR InterPro; IPR001956; CBD 3.  
DR Pfam; PF00942; CEM 3; 1.  
DR PRODOM; PD001947; CBD 3; 1.  
DR Cellulose degradation; Hydrolase; Glycosidase; Signal.  
FT SIGNAL 1 33 POTENTIAL.  
FT CHAIN 34 700 ENDOGLUCANASE A.  
FT DOMAIN 548 700 CELLULOSE-BINDING (BY SIMILARITY).  
FT ACT SITE 213 213 BY SIMILARITY.  
SQ SEQUENCE 700 AA; 76910 MW; 3D5C8CADAS3EEB0F CRC64;  
  
Query Match 36.6%; Score 172; DB 1; Length 700;  
Best Local Similarity 43.9%; Pred. No. 3.4e-13;  
Matches 36; Conservative 18; Mismatches 25; Indels 3; Gaps 2;

DR HSSP; Q06851; INBC.  
DR InterPro; IPR001956; CBD 3.  
DR InterPro; IPR008965; Cellul\_bind.  
DR Pfam; PF00942; CEM 3; 1.  
DR PRODOM; PD001947; CBD 3; 1.  
KW Hypothetical protein.  
FT NON\_TER 1  
SQ SEQUENCE 145 AA; 15782 MW; 9514E3A71B106AEB CRC64;  
  
Query Match 38.4%; Score 180.5; DB 1; Length 145;  
Best Local Similarity 43.9%; Pred. No. 3.4e-13;  
Matches 36; Conservative 18; Mismatches 25; Indels 3; Gaps 2;  
  
QY 7 VOYKNDSPAGDNIKPGQLVNTGSSVDLSVTYVYFTRDGGSTLVNCDWAAMGC 66  
DB 1 LQYRAADTNAADNQIKPSFNKNNGTSAVDLSLKIRYFTKDGSAVNGW-IDWAQLGG 59  
  
QY 67 GNIRASFGSNVPATPTADTYLQ 88  
DB 60 SNIQISFG--NHTGTNSDTYVE 79  
  
RESULT 8  
GUNA\_PAEIA STANDARD; PRT; 700 AA.  
ID GUNA\_PAEIA  
AC P29719;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)  
DE (Cellulase A) (EG-A).  
GN CELA.  
OS Paenibacillus lautus (Bacillus lautus).  
OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.  
OX NCBI\_TaxID=1401;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PL236;  
RX MEDLINE=92276330; PubMed=1592807;  
RA Hansen C.K., Diderichsen B., Joergensen P.L.;  
RT "celsA from Bacillus lautus PL236 encodes a novel cellulose-binding  
endo-beta-1,4-glucanase."  
RL J. Bacteriol. 174:3522-3531(1992).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
-----  
DR EMBL; M76588; AAA22303.1; .  
DR PIR; B41897; B41897.  
DR HSSP; Q06851; INBC.  
DR InterPro; IPR001956; CBD 3.  
DR Pfam; PF00942; CEM 3; 1.  
DR PRODOM; PD001947; CBD 3; 1.  
DR Cellulose degradation; Hydrolase; Glycosidase; Signal.  
FT SIGNAL 1 33 POTENTIAL.  
FT CHAIN 34 700 ENDOGLUCANASE A.  
FT DOMAIN 548 700 CELLULOSE-BINDING (BY SIMILARITY).  
FT ACT SITE 213 213 BY SIMILARITY.  
SQ SEQUENCE 700 AA; 76910 MW; 3D5C8CADAS3EEB0F CRC64;  
  
Query Match 36.6%; Score 172; DB 1; Length 700;  
Best Local Similarity 43.9%; Pred. No. 3.4e-13;  
Matches 36; Conservative 18; Mismatches 25; Indels 3; Gaps 2;

	Best Local Similarity	44.9%, Pred. No. 2e-ll;
	Matches	Conservative 18; Mismatches 27; Indels 4; Gaps 3;
Qy	1 VSGGVKQQYKNNO--SAPDGNQIKPGLQLVNTGSSVDLSTVTVRHYWTRDGSGSLTVNC	59
	:       :	:         :
Dd	549 VNSDLVVQYKDGRNNAATDNQIKPHFNQKNGTSVPDLSSLTRYPTKD--SSAAMGW	607
Qy	60 DWAAAGCGNIRASFGSVNPATPTADTYLQ	88
Dd	608 DWAKLGGSNIQISPGNHGA--SDSYAE	634
	:	:

RESULT 9	GUNI_CLOTHM	STANDARD;	PRT;	879 AA.
ID	GUNI_CLOTHM	STANDARD;	PRT;	879 AA.
AC	Q02934;			
DT	01-FEB-1995 (Rel. 31, Created)			
DT	01-FEB-1995 (Rel. 31, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Endoglucanase I precursor (EC 3.2.1.4) (EGI) (Endo-1,4-beta-glucanase) (Cellulase I).			
GN	CELL.			
OS	Clostridium thermocellum.			
OS	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;			
OC	Clostridium.			
OX	NCBI_TaxID=1515;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 56-69.			
RC	STRAIN=NCIB 10682;			
RX	MEDLINE=93171873; PubMed=8436949;			
RA	Hazlewood G.P., Davidson K., Laurie J.I., Huskisson N.S.,			
RA	Gilbert H.J.;			
RT	"gene sequence and properties of CellI, a family E endoglucanase from			
RT	Clostridium thermocellum.";			
RL	J. Gen. Microbiol. 139:307-316(1993).			
CC	-!- FUNCTION: THIS ENZYME CATALYZES THE ENDOHYDROLYSIS OF 1,4-BETA-			
CC	GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-			
CC	GLUCANS. PRINCIPALLY ACTIVE AGAINST BARLEY BETA-GLUCAN.			
CC	-!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic			
CC	linkages in cellulose, lichenin and cereal beta-D-glucans.			
CC	-!- PATHWAY: Cellulose degradation.			
CC	-!- SIMILARITY: Belongs to cellulase family E (family 9 of glycosyl			
CC	hydrolases).			

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL, outstation of the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))

[illegible]

Query Match	35.2%	Score 165.5;	DB 1;	Length 879;
Best Local Similarity	33.0%	Pred. No. 1.4e-10;		
Matches	29;	Conservative 26;	Mismatches 32;	Gaps 1;
Qy	1	VSGGVKVVQYKNNDSPGDNQIKPGLQLVNTGSSVDLSTVTWYWRTRDGGSSSTLVNCD	60	
Db	736	IKGEVLTVLYANGNAGATSNSINPRFKIINGTKAINLSVKIRYYVTKEGGASQNF-W	794	
Qy	61	WAAMCGCNIIRASTGSVNPAFTPTADTVLQ	88	
Db	795	WSGAGNSVNTGNFNLSSPKEGADTCL	822	

```

RESULT 10
CIPB_CLOTM STANDARD; PRT; 772 AA.
ID CIPB_CLOTM STANDARD; PRT; 772 AA.
AC Q01866;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cellulosomal scaffolding protein B (Cellulosomal glycoprotein sl/SL)
DE (Cellulose integrating protein B) (Fragment).
GN CtpB.
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1515;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YS;
RX MEDLINE=93146373; PubMed=1490597;
RA Poole D.M., Morag E., Lamed R., Bayer E.A., Hazlewood G.P.,
RA Gilbert H.J.;
RT "Identification of the cellulose-binding domain of the cellulosome
RL FEMS Microbiol. Lett. 78:181-186(1992).
CC -1- FUNCTION: ACTS AS A SCAFFOLDING PROTEIN IN THE CELLULOSE. IT
CC PROMOTES BINDING OF CELLULOSE TO THE CATALYTIC DOMAINS OF THE
CC CELLULOLYTIC ENZYMES PROBABLY THROUGH THE BINDING OF THE NINE
CC REPEATED DOMAINS WITH THE DUPLICATED SEGMENT OF 24 AMINO ACIDS
CC PRESENT IN CATALYTIC SUBUNITS OF THE CELLULOSE.
CC -1- SUBCELLULAR LOCATION: Cell surface.
CC -1- DOMAIN: THE COHESIN DOMAINS BIND TO THE DOCKERIN DOMAIN BORN BY
CC THE CATALYTIC COMPONENTS OF THE CELLULOSE.
CC -1- SIMILARITY: Contains at least 3 cohesin domains.
CC -1- SIMILARITY: Contains 2 dockerin domains.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
-----
ENBL; X68233; CAA48312.1; -.
DR HSSP; Q05851; INEC
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR002102; Cohesin.
DR InterPro; IPR002105; Dockerin_1.
DR InterPro; IPR002048; EF-hand.
DR Pfam; PF00942; CEM_3; 1.
DR Pfam; PF00963; Cohesin; 3.
DR Pfam; PF00404; Dockerin_1.
DR ProDom; PD001947; CBD_3; 1.
DR PROSITE; PS00018; EF HAND; UNKNOWN 1.
DR PROSITE; PS00448; CL05_CELLULOSE_RPT; 2.
KW Cellulose degradation; Cell wall; Glycoprotein; Repeat.
FT NON TER 1 1
FT DOMAIN <1 80 COHESIN 1.
FT DOMAIN 81 93 LINKER (PRO/THR-RICH).
FT

```

```

FT DOMAIN          94 240 COHESIN 2.
FT DOMAIN          241 272 LINKER (PRO/THR-RICH).
FT DOMAIN          273 439 CELLULOSE-BINDING.
FT DOMAIN          440 461 LINKER (PRO/THR-RICH).
FT DOMAIN          462 607 COHESIN 3.
FT DOMAIN          710 733 DOCKERIN 1.
FT DOMAIN          743 766 DOCKERIN 2.
SQ SEQUENCE       772 AA; 83491 MW; BBF06DE5E094FE10 CRC64;

Query Match          35.1%; Score 165; DB 1; Length 772;
Best Local Similarity 36.1%; Pred. No. 1.4e-10;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2

QY 1 VSGGVKVVQKNNDSPEDNQIKPLGLVNTGSSVDLSTVTRVYWFTRDGGSSTLVYNCD 60
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 277 VSGNLKVFYNPSDDTNSINPQKVTNTGSSAIDLKSLRLYYTYVDGQKDTFM-CD 335
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 61 WAAM-----GCGNIRASFGSVNPATPTADTYLQ 88
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 336 HAAIIGSGSYNGITSNVKGTFVKMSSSTNNADTYLE 372
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 11
CIPA_CLOTH
ID CIPA_CLOTH STANDARD; PRT; 1853 AA.
AC Q06851;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cellulosomal scaffolding protein A precursor (Cellulosomal
DE glycoprotein SI/SL) (Cellulose integrating protein A) (Cohesin).
GN CIPA.
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1515;
RX [1]
RX SEQUENCE FROM N.A., AND SEQUENCE OF 29-40 AND 42-43.
RX STRAIN=ATCC 27405 / DSM 1237;
RX MEDLINE=93302508; PubMed=9316083;
RX Geringross U.T., Romaniec M.P.M., Kobayashi T., Huskisson N.S.,
RX Demain A.L.;
RA "Sequencing of a Clostridium thermocellum gene (cipA) encoding the
RT cellulosomal SI-protein reveals an unusual degree of internal
RT homology.";
RL Mol. Microbiol. 8:325-334(1993).
RN [2]
RN SEQUENCE OF 1820-1853 FROM N.A.
RX MEDLINE=93209931; PubMed=8458832;
RX Fujino T., Beguin P., Aubert J.-P.;
RA "Organization of a Clostridium thermocellum gene cluster encoding the
RT cellulosomal scaffolding protein Cipa and a protein possibly involved
RT in attachment of the cellulosome to the cell surface.";
RL J. Bacteriol. 175:1891-1899(1993).
RN [3]
RN X-RAY CRYSTALLOGRAPHY (2.15 ANGSTROMS) OF 184-321.
RX MEDLINE=97238934; PubMed=9083107;
RA Shimon L.J.W., Bayer E.A., Morag E., Lamed R., Yaron S., Shoham Y.,
RA Frolow F.;
RA "A cohesin domain from Clostridium thermocellum: the crystal
RT structure provides new insights into cellulosome assembly.";
RL Structure 5:381-390(1997).
RN [4]
RN X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 368-522.
RX MEDLINE=97076134; PubMed=8918451;
RA Tormo J., Lamed R., Chirino A.J., Morag E., Bayer E.A., Shoham Y.,
RA Steitz T.A.;
RA "Crystal structure of a bacterial family-III cellulose-binding
RT domain: a general mechanism for attachment to cellulose.";
RL EMBO J. 15:5739-5751(1996).
RN [5]
RN X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 1213-1361.
RX MEDLINE=98022914; PubMed=9402065;

```





Search completed: May 11, 2004, 12:07:29  
Job time : 4.42639 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 11:45:55 ; Search time 10.385 Seconds  
(without alignments)  
2704.020 Million cell updates/sec

Title: US-09-917-376-4

Perfect score: 470

Sequence: 1 VSGGVKQYKNDSAPGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_25:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phase:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_rvirus:\*

16: sp\_bacteriaph:\*

17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	257	54.7	741	16 Q82QF2	Q82qf2 streptomyc
2	228.5	48.6	616	2 Q7X2N2	Q7x2n2 thermomon
3	225.5	48.0	170	2 Q9RFK6	Q9rfx6 caldibacill
4	216.5	45.1	930	2 Q9RFK5	Q9rfx5 caldibacill
5	214.5	45.6	921	2 Q9L8L8	Q9l8l8 caldibacill
6	209.5	44.6	1091	2 Q8KKF7	Q8kkf7 paenibacill
7	204.5	43.5	997	2 Q9Z4I1	Q9z4i1 bacillus sp
8	194.5	41.4	1751	2 Q9AQA4	Q9agg4 caldicellul
9	192.5	41.0	1000	2 Q24820	Q24820 thermophili
10	192.5	41.0	1770	2 Q9X3P5	Q9x3p5 caldicellul
11	191.5	40.7	261	2 Q9AQG7	Q9agg7 caldicellul
12	191.5	40.7	1426	2 Q9X3P6	Q9x3p6 caldicellul
13	191	40.6	1711	2 P96311	P96311 anaerocellul
14	190.5	40.5	996	2 Q9AQH0	Q9aqh0 caldicellul
15	190.5	40.5	1779	2 O52374	O52374 caldicellul
16	190	40.4	499	2 Q93TJ6	Q93tj6 bacillus su

17	190	40.4	508	2	Q93LD0	Q93ld0 bacillus su
18	187	39.8	499	2	O52731	O52731 bacillus sp
19	186	39.6	499	2	O45532	O45532 bacillus su
20	186	39.6	501	2	O83012	O83012 bacillus sp
21	180	38.3	486	2	O45430	O45430 bacillus sp
22	177	37.7	499	2	O8RPQ6	O8rpq6 bacillus am
23	172.5	36.7	1915	2	Q9RPL0	Q9rp0 acetivibrio
24	168	35.7	473	16	Q9RK75	Q9rk75 streptomyc
25	165.5	35.2	887	2	Q9L3J8	Q9l3j8 clostridium
26	157	33.4	542	2	Q7X3S6	Q7x3s6 bacillus li
27	150.5	32.0	2316	2	Q9FDJ9	Q9fdj9 bacteroides
28	130.5	27.8	1483	16	Q977Y4	Q977y4 clostridium
29	130.5	27.8	1546	2	Q45996	Q45996 clostridium
30	126.5	26.9	1162	2	O82830	O82830 clostridium
31	119.5	25.4	307	2	Q46392	Q46392 clostridium
32	117	24.9	1230	2	O59325	O59325 clostridium
33	107	22.8	221	2	O8VVI7	O8vvi7 clostridium
34	87	18.5	440	2	Q93LI9	Q93li9 bacillus sp
35	78	16.6	135	13	Q90542	Q90542 ginglymosto
36	76	16.2	121	13	O8JGA5	O8jga5 ginglymosto
37	76	16.2	121	13	O8JGA9	O8jga9 ginglymosto
38	75.5	16.1	618	4	Q9Y3Z2	Q9y3z2 homo sapien
39	75.5	16.1	660	2	Q9L3J2	Q9l3j2 clostridium
40	75.5	16.1	986	4	O94858	O94858 homo sapien
41	75.5	16.1	1300	4	Q8IZF2	Q8izf2 homo sapien
42	73.5	15.6	547	5	P91006	P91006 caenorhabdi
43	73.5	15.6	1428	16	O8YRU7	O8yru7 anabaena sp
44	73	15.5	511	5	Q9GYG5	Q9gyg5 caenorhabdi
45	72	15.3	499	12	Q993M3	Q993m3 autonomous

ALIGNMENTS

RESULT 1

ID	Q82QF2	PRELIMINARY;	PRT;	741 AA.
AC	Q82QF2;			
DT	01-JUN-2003 (TREMELrel. 24, Created)			
DT	01-JUN-2003 (TREMELrel. 24, Last sequence update)			
DT	01-OCT-2003 (TREMELrel. 25, Last annotation update)			
DE	Putative cellulose 1,4-beta-cellobiosidase.			
GN	GUXA1 OR SAV557.			
OS	Streptomyces avermitilis.			
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC	Streptomycineae; Streptomycetaceae; Streptomyces.			
OX	NCBI_TaxID=33903;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;			
RC	MEDLINE=21477403; PubMed=11572948;			
RA	Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,			
RA	Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,			
RA	Kikuchi H., Shiba T., Sakaki Y., Hattori M.;			
RT	"Genome sequence of an industrial microorganism Streptomyces			
RT	avermitilis: deducing the ability of producing secondary			
RT	metabolites.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;			
RC	MEDLINE=22608306; PubMed=12692562;			
RA	Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,			
RA	Sakaki Y., Hattori M., Omura S.;			
RT	"Complete genome sequence and comparative analysis of the industrial			
RT	microorganism Streptomyces avermitilis.";			
RL	Nat. Biotechnol. 21:526-531(2003).			
EMBL	AF005023; BAC68267.1; -			
DR	GO; GO:0004553; F:Hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.			
DR	GO; GO:0005975; P:carbohydrate metabolism; IEA.			
DR	InterPro; IPR001956; CBD 3.			
DR	InterPro; IPR008965; Cellul_bind.			
DR	InterPro; IPR003961; FN_III.			



```
DR InterPro; IPR008957; FN.III-like.
DR InterPro; IPR001524; Glyco_hydro_6.
DR Pfam; PF00942; CBM_3; 1.
DR Pfam; PF00441; fn3; 1.
DR Pfam; PF01341; Glyco_hydro_6; 1.
DR PRINTS; PR00733; GLHYDRLAS.E6.
DR ProDom; PD001947; CBD_3; 1.
DR ProDom; PD003733; Glyco_hydro_6; 1.
DR SMART; SM0060; FN3; 1.
DR PROSITE; PS00655; GLYCOSYL_HYDROL_F6_1; 1.
DR Complete proteome.
SQ SEQUENCE 741 AA; 77396 MW; 79404B40B2B4A7AF CRC64;

Query Match 54.7%; Score 257; DB 16; Length 741;
Best Local Similarity 56.3%; Pred. No. 1.7e-19;
Matches 49; Conservative 13; Mismatches 25; Indels 0; Gaps 0;

QY 2 SGGVKQVYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRWFTRDGGSSTLVNCWDW 61
DB 591 SGLKLVYKNNDSATDNIAPGLRIVNTGSGSLDSKVYRFTYRDSGSPVNAWCYD 650

QY 62 AAMGCGNIRASFGSVNPAFTADTYLQ 88
DB 651 AAVGCSNVSLKVPLTTPVPGADAYLE 677

RESULT 2
ID Q7X2N2 PRELIMINARY; PRT; 616 AA.
AC Q7X2N2;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Endoglucanase.
GN CEL5B.
OS Thermomonospora fusca.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptoporangineae; Nocardiopsaceae; Thermobifida.
OX NCBI_TaxID=2021;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TW51;
RA Posta K., Beki E., Kukolya J., Hornok L.;
RT "Phylogenetic relationships of Tfi cel5B, a new endoglucanase encoding
RT gene from Thermobifida fusca.";
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY298814; AAP56348.1; --
SQ SEQUENCE 616 AA; 67701 MW; 24FFC1EA1A3F5639 CRC64;

Query Match 48.6%; Score 228.5; DB 2; Length 616;
Best Local Similarity 52.3%; Pred. No. 1.8e-16;
Matches 45; Conservative 16; Mismatches 22; Indels 3; Gaps 2;

QY 2 SGGVKQVYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRWFTRDGGSSTLVNCWDW 61
DB 471 TGALEYVYRNLSAADSQTAPGLRLVNTGSSVDLADVEIHYFTNEPG-GTLQFTCDW 529

QY 62 AAMGCGNIRASFGSVNPAFTADTYL 87
DB 530 AQVGCANVASFTSL--SAPGADTSL 553

RESULT 3
ID Q9RFK6 PRELIMINARY; PRT; 170 AA.
AC Q9RFK6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
OS Caldicellulosus cellulovorans.
OC Bacteria; Firmicutes; Bacillales; Alicyclobacillaceae; Caldicellulosus.
OX NCBI_TaxID=74586;
```

```

[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=20120520; PubMed=10653733;
RA Sunna A., Gibbs M.D., Chin C.W.J., Nelson P.J., Bergquist P.L.;
RT "A gene encoding a novel multidomain beta-1,4-mannanase from
RT Caldicellulosus cellulovorans and action of the recombinant enzyme on
RT kraft pulp.";
RL Appl. Environ. Microbiol. 66:664-670(2000).
DR EMBL; AF163837; AAF22273.1; --
DR HSP; Q06851; INBC.
GO: GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul_bind.
DR Pfam; PF00942; CBM_3; 1.
DR ProDom; PD001947; CBD_3; 1.
KW Hypothetical protein.
FT NON TER 1
SQ SEQUENCE 170 AA; 16493 MW; 7AC9D33F44E3A0B4 CRC64;

Query Match 48.0%; Score 225.5; DB 2; Length 170;
Best Local Similarity 47.7%; Pred. No. 8.2e-17;
Matches 41; Conservative 17; Mismatches 27; Indels 1; Gaps 1;

QY 3 GGVKQVYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRWFTRDGGSSTLVNCWDW 62
DB 20 GSLVQVYRAADTNAGDNQKPHFRIVNRGTSSVPLSELTIYTYTVD-GDKPQVFNCDWA 78.

QY 63 AAMGCGNIRASFGSVNPAFTADTYLQ 88
DB 79 QVGCNVRGSGFVKLSTORTGADYYIE 104

RESULT 4
ID Q9RFK5 PRELIMINARY; PRT; 930 AA.
AC Q9RFK5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Multidomain beta-1,4-mannanase precursor.
GN MANA.
OS Caldicellulosus cellulovorans.
OC Bacteria; Firmicutes; Bacillales; Alicyclobacillaceae; Caldicellulosus.
OX NCBI_TaxID=74586;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20120520; PubMed=10653733;
RA Sunna A., Gibbs M.D., Chin C.W.J., Nelson P.J., Bergquist P.L.;
RT "A gene encoding a novel multidomain beta-1,4-mannanase from
RT Caldicellulosus cellulovorans and action of the recombinant enzyme on
RT kraft pulp.";
RL Appl. Environ. Microbiol. 66:664-670(2000).
DR EMBL; AF163837; AAF22274.1; --
DR HSP; Q06851; INBC.
GO: GO:0010928; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
GO: GO:0004553; P:carbohydrate metabolism; IEA.
GO: GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR004302; Chitin_binding_3.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00942; CBM_3; 2.
DR Pfam; PF00150; cellulase; 1.
DR Pfam; PF03087; Chitin_bind_3; 1.
DR ProDom; PD001947; CBD_3; 2.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW SIGNAL.
FT SIGNAL 1 33 POTENTIAL.
FT CHAIN 34 930 MULTIDOMAIN BETA-1,4-MANNANASE.
SQ SEQUENCE 930 AA; 101576 MW; 0086638D54DIA2CC CRC64;

Query Match 46.1%; Score 216.5; DB 2; Length 930;
Best Local Similarity 46.5%; Pred. No. 6.2e-15;
```

Cellulose 1,4-beta-cellobiosidase precursor (EC 3.2.1.91).

CEL48C.

Paenibacillus sp. BP-23.

Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.

NCBI\_TaxID=198119;

[1]

SEQUENCE FROM N.A.

STRAIN=BP-23;

RA Sanchez M.M., Pastor F.I.J., Diaz P.;

RT "Paenibacillus sp. BP-23 family 48 cellulase. Cloning and performance

on cellulosic substrates";

RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.

EMBL; AJ48933; CAD32945.1; -

DR GO; GO:0016162; F:cellulose 1,4-beta-cellobiosidase activity; IEA.

DR GO; GO:0005975; P:carbohydrate metabolism; IEA.

DR InterPro; IPR001956; CBD 3.

DR InterPro; IPR008965; Cellul\_bind.

DR InterPro; IPR003961; FN\_III.

DR InterPro; IPR008957; FN\_III-like.

DR InterPro; IPR000556; Glyco\_hydro\_48.

DR InterPro; IPR008928; Glyco\_trans\_6hp.

DR Pfam; PF00942; CEM 3; 1.

DR Pfam; PF00041; fn3; 1.

DR Pfam; PF02011; Glyco\_hydro\_48; 1.

DR PRINTS; PR00844; GLHYDRLASE48.

DR ProDom; PD001947; CBD 3; 1.

DR ProDom; PD011903; Glyco\_hydro\_48; 1.

DR SMART; SM00060; FN3; 2.

KW Signal; Hydrolase; Glycosidase.

FT SIGNAL 1 35 POTENTIAL.

FT CHAIN 36 1091 CELLULOSE 1,4-BETA-CELLOBIOSIDASE.

QQ SEQUENCE 1091 AA; 118001 MW; 21EACCEB2E70478 CRC64;

Query Match 44.6%; Score 209.5; DB 2; Length 1091;

Best Local Similarity 46.0%; Pred. No. 4.4e-14;

Matches 40; Conservative 21; Mismatches 25; Indels 1; Gaps 1;

Qy 2 SGGVKQVQKNDSPAGDNOIKPGLQLVNTGSSVDLSTVTVRVWFTRDGGSTLVNCDW 61

Db 940 TGILEVQVRNGSGASGNALTPQFNLKNTGTTAIDLSKVVRVYFTKD-SAADM5FWCDY 998

Qy 62 AAMCGCNIRASFSGSVNPATPTADTYLQ 88

Db 999 AOLGSANVQGSFVAVNPAPKGTADTYLE 1025

RESULT 7

Q92411 PRELIMINARY; PRT; 997 AA.

ID Q92411

AC Q92411;

DT 01-MAY-1999 (TrEMBLrel. 10, Created)

DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Cellulase precursor (EC 3.2.1.4).

GN CELB.

OS Bacillus sp. BP-23

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

NCBI\_TaxID=89769;

[1]

SEQUENCE FROM N.A.

STRAIN=BP-23;

RC MEDLINE=21129642; PubMed=11234960;

RX Pastor F.I.J., Pujol X., Blanco A., Vidal T., Torres A.L., Diaz P.;

RA "Molecular cloning and characterization of a multidomain endoglucanase

from Paenibacillus sp BP-23: evaluation of its performance in pulp

refining";

RT Appl. Microbiol. Biotechnol. 55:61-68(2001).

RL -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.

CC EMBL; AJ133614; CAB38941.1; -

DR HSP; P26221; IFF4.

DR GO; GO:0008810; F:cellulase activity; IEA.

DR GO; GO:0005215; F:transporter activity; IEA.

```
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR008965; Cellul bind.
DR InterPro; IPR003961; FN III-like.
DR InterPro; IPR008957; FN III-like.
DR InterPro; IPR001701; Glyco_hydro_9.
DR InterPro; IPR008928; Glyco_trans_6hp.
DR InterPro; IPR000566; LipocIn_cytFABP.
DR Pfam; PF00942; CBM 3; 2.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF00759; Glyco_hydro_9; 1.
DR ProDom; PD001947; CBD_3; 1.
DR SMART; SM00060; FN3; 2.
DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.
DR PROSITE; PS00213; LIPOCALIN; 1.
KW Glycosidase; Hydrolase; Signal.
FT SIGNAL 1 34 POTENTIAL.
SQ SEQUENCE 997 AA; 106927 MW; F20CB119D6410629 CRC64;

Query Match 43.5%; Score 204.5; DB 2; Length 997;
Best Local Similarity 43.7%; Pred. No. 1.4e-13;
Matches 38; Conservative 21; Mismatches 27; Indels 1; Gaps 1;

QY 2 SGGVQVQKNDSPAGDQNIKPGQLQVNTGSSVDLSVTYVYFTRDGGSSTLVYNCW 61
Db 846 TCTLEVQVRSGGSGNSNAVTQFNLKNTGTQAIIDLSTVKIRYFTKD-GTELSFWC 904

QY 62 AMWGCNIRASFGSVNTPATPTADTYLQ 88
Db 905 AQVGSANVQGMFVAVNPARGTADTYVE 931

RESULT 8
Q9A0G4 PRELIMINARY; PRT; 1751 AA.
AC Q9A0G4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Cell.
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20171169; PubMed=10706665;
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P., Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL Curr. Microbiol. 40:333-340(2000).
DR EMBL; AF078042; AAK06394.1; -.
DR HSSP; Q06851; INBC.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0002215; F:transporter activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR008965; Cellul bind.
DR InterPro; IPR001701; Glyco_hydro_9.
DR InterPro; IPR008928; Glyco_trans_6hp.
DR InterPro; IPR000566; LipocIn_cytFABP.
DR Pfam; PF00942; CBM 3; 4.
DR Pfam; PF00759; Glyco_hydro_9; 1.
DR ProDom; PD001947; CBD_3; 3.
DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.
DR PROSITE; PS00213; LIPOCALIN; 1.
SQ SEQUENCE 1751 AA; 192176 MW; 60178CBF3C00BE95 CRC64;

Query Match 41.0%; Score 192.5; DB 2; Length 1000;
Best Local Similarity 43.7%; Pred. No. 2.9e-12;
Matches 38; Conservative 17; Mismatches 31; Indels 1; Gaps 1;

QY 2 SGGVQVQKNDSPAGDQNIKPGQLQVNTGSSVDLSVTYVYFTRDGGSSTLVYNCW 61
Db 372 SGQIKVLYANKETNSTTIRPWLKVNTGSSSIDLSRVTIRYWTVDGDKAQSASV-DW 430

QY 62 AMWGCNIRASFGSVNTPATPTADTYLQ 88
Db 431 AQIGASNVTFKFKVLSSSVSGADTYLE 457

RESULT 10
Q9X3P5 PRELIMINARY; PRT; 1770 AA.
AC Q9X3P5;
```

```
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE XYN.
GN XYN.
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tok7B.1;
RX MEDLINE=20171169; PubMed=10706665;
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL Curr. Microbiol. 40:333-340(2000).
DR EMBL; AF078040; AAK06391.1; -.
DR HSSP; Q06851; INBC.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul_bind.
DR Pfam; PF00942; CBM_3; 1.
DR ProDom; PD001947; CBD_3; 1.
DR ProDom; PD001947; CBD_3; 1.
KW Hydrolase.
FT NON_TER 261 261
FT NON_TER 261 261
SQ SEQUENCE 261 AA; 28759 MW; 4771744A26A6AE04 CRC64;

Query Match 40.7%; Score 191.5; DB 2; Length 261;
Best Local Similarity 43.5%; Pred. No. 7.3e-13;
Matches 37; Conservative 15; Mismatches 32; Indels 1; Gaps 1;

QY 4 GVKVQYKNDSPAGDNQIKPGLQVNTGSSVDLSVTYVRYWFTRDGSSSTLVYNCDAWAA 63
DB 118 GLKVLKNNETSASTSIRPFKIVNGSSVDLSRVKIRYWTVDGKPKQSAV-CDWAQ 176
QY 64 MCGNIRASFSGSVNPATPTADTYLQ 88
DB 177 IGASNVTFNFVKLSGSGVADYYLE 201

RESULT 12
O9X3P6 PRELIMINARY; PRT; 1426 AA.
AC O9X3P6;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE CelB.
GN CELB.
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tok7B.1;
RX MEDLINE=20171169; PubMed=10706665;
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL Curr. Microbiol. 40:333-340(2000).
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY F (FAMILY 10 OF GLYCOSYL
HYDROLASES).
DR EMBL; AF078737; AAD30364.1; -.
DR HSSP; Q06851; INBC.
DR GO; GO:0003779; F:actin binding; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001589; Actbind_actnin.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR001000; Glyco_hydro_10.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00942; CBM_3; 3.
DR Pfam; PF00150; cellulase; 1.
DR Pfam; PF00331; Glyco_hydro_10; 1.
DR PRINTS; PR00134; GLHYDRLASE10.
DR ProDom; PD001947; CBD_3; 3.
DR SMART; SM00633; Glyco_10; 1.
DR PROSITE; PS00019; ACTININ 1; 1.
DR PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.
SQ SEQUENCE 1770 AA; 193641 MW; 8BAF1937D4926C92 CRC64;

Query Match 41.0%; Score 192.5; DB 2; Length 1770;
Best Local Similarity 43.5%; Pred. No. 5.7e-12;
Matches 37; Conservative 16; Mismatches 31; Indels 1; Gaps 1;

QY 4 GVKVQYKNDSPAGDNQIKPGLQVNTGSSVDLSVTYVRYWFTRDGSSSTLVYNCDAWAA 63
DB 1104 GLKVLKNNETSASTSIRPFKIVNGSSVDLSRVKIRYWTVDGKPKQSAV-CDWAQ 1162
QY 64 MCGNIRASFSGSVNPATPTADTYLQ 88
DB 1163 IGASNVTFNFVKLSGSGVADYYLE 1187

RESULT 11
O9AQG7 PRELIMINARY; PRT; 261 AA.
AC O9AQG7;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Glycosyl hydrolase 6 (Fragment).
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tok7B.1;
RX MEDLINE=20171169; PubMed=10706665;
```

```
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL Curr. Microbiol. 40:333-340(2000).
DR EMBL; AF078040; AAK06391.1; -.
DR HSSP; Q06851; INBC.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul_bind.
DR Pfam; PF00942; CBM_3; 1.
DR ProDom; PD001947; CBD_3; 1.
DR ProDom; PD001947; CBD_3; 1.
KW Hydrolase.
FT NON_TER 261 261
FT NON_TER 261 261
SQ SEQUENCE 261 AA; 28759 MW; 4771744A26A6AE04 CRC64;

Query Match 40.7%; Score 191.5; DB 2; Length 261;
Best Local Similarity 43.5%; Pred. No. 7.3e-13;
Matches 37; Conservative 15; Mismatches 32; Indels 1; Gaps 1;

QY 4 GVKVQYKNDSPAGDNQIKPGLQVNTGSSVDLSVTYVRYWFTRDGSSSTLVYNCDAWAA 63
DB 118 GLKVLKNNETSASTSIRPFKIVNGSSVDLSRVKIRYWTVDGKPKQSAV-CDWAQ 176
QY 64 MCGNIRASFSGSVNPATPTADTYLQ 88
DB 177 IGASNVTFNFVKLSGSGVADYYLE 201

RESULT 12
O9X3P6 PRELIMINARY; PRT; 1426 AA.
AC O9X3P6;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE CelB.
GN CELB.
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tok7B.1;
RX MEDLINE=20171169; PubMed=10706665;
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL Curr. Microbiol. 40:333-340(2000).
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY F (FAMILY 10 OF GLYCOSYL
HYDROLASES).
DR EMBL; AF078737; AAD30364.1; -.
DR HSSP; Q06851; INBC.
DR GO; GO:0003779; F:actin binding; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001589; Actbind_actnin.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR001000; Glyco_hydro_10.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00942; CBM_3; 3.
DR Pfam; PF00150; cellulase; 1.
DR Pfam; PF00331; Glyco_hydro_10; 1.
DR PRINTS; PR00134; GLHYDRLASE10.
DR ProDom; PD001947; CBD_3; 3.
DR SMART; SM00633; Glyco_10; 1.
DR PROSITE; PS00019; ACTININ 1; 1.
DR PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.
SQ SEQUENCE 1770 AA; 193641 MW; 8BAF1937D4926C92 CRC64;
```

```
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 1426 AA; 157544 MW; 29B3FDB85D09A863 CRC64;

Query Match
Best Local Similarity 40.7%; Score 191.5; DB 2; Length 1426;
Matches 37; Conservative 15; Mismatches 32; Indels 1; Gaps 1;

QY 4 GVKVOYKNNDSAPGDNQIKPGQLVNTGSSVDLSVTYVRYWFTRDGSGSTLVYNCDAWAA 63
Db 413 GLKVLKNNETSASTGSIREFWFKVNGSSVDLSRVKIRYWTVDGDKPQSAV-CDWAAQ 471
QY 64 MCGCNIRASFGSVNPTATPTADTYLQ 88
Db 472 IGASNVTFNFVKLSGSGVADYYLE 496

RESULT 13
P96311 PRELIMINARY; PRT; 1711 AA.
AC P96311;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE 1,4-beta-glucanase (fragment).
GN CELA.
OS Anaerocellum thermophilum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Anaerocellum group;
OC Anaerocellum.
OX NCBI_TaxID=31899;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z-1320;
RX MEDLINE=98154434; PubMed=9493383;
RA Zverlov V., Mahr S., Riedel K., Bronnenmeier K.;
RT "Properties and gene structure of a bifunctional cellulolytic enzyme (Cela) from the extreme thermophile Anaerocellum thermophilum with separate glycosyl hydrolase family 9 and 48 catalytic domains.";
RL Microbiology 144:457-465(1998).
DR EMBL; Z86105; CAB06786.1; -.
DR PIR; T31337; T31337.
DR HSSP; P26221; 1TF4.
DR GO; GO:0004553; P:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR008965; Cellul. bind.
DR InterPro; IPR000556; Glyco_hydro_48.
DR InterPro; IPR001701; Glyco_hydro_9.
DR InterPro; IPR008928; Glyco_trans_6hp.
DR Pfam; PF00942; CBM 3; 3.
DR Pfam; PF02011; Glyco_hydro_48; 1.
DR Pfam; PF00759; Glyco_hydro_9; 1.
DR PRINTS; PR00844; GLHYDRLASE48.
DR ProDom; PD001947; CBD 3; 2.
DR ProDom; PD011903; Glyco_hydro_48; 1.
DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.
FT NON TER 1
SQ SEQUENCE 1711 AA; 189979 MW; E3E987CBB9CD0C21 CRC64;

Query Match
Best Local Similarity 40.6%; Score 191; DB 2; Length 1711;
Matches 39; Conservative 18; Mismatches 30; Indels 2; Gaps 2;

QY 1 VSGG-VKVOYKNNDSAPGDNQIKPGQLVNTGSSVDLSVTYVRYWFTRDGSGSTLVYNC 59
Db 683 VAGGOIKVLKYANKETNSTTIRPWLKVVNTGSSVDLSRVKIRYWTVDGDKPQSAIS- 741
QY 60 DWAAAGCGNIRASFGSVNPTATPTADTYLQ 88
Db 742 DWAGIAGSNVTFKFKVLSGSGVADYYLE 770

RESULT 14
Q9AQH0 PRELIMINARY; PRT; 996 AA.
AC Q9AQH0;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Glycosyl hydrolase 5 (fragment).
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tok7B.1;
RX MEDLINE=20171169; PubMed=10706665;
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P., Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL Curr. Microbiol. 40:333-340(2000).
DR EMBL; AF078038; AAK06388.1; -.
DR HSSP; Q06851; INBC.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR008965; Cellul. bind.
DR InterPro; IPR002860; GH_BNR.
DR Pfam; PF02012; BNR; 9.
DR Pfam; PF00942; CBM 3; 1.
DR ProDom; PD001947; CBD 3; 1.
DR KW Hydrolase.
FT NON TER 996
SQ SEQUENCE 996 AA; 108275 MW; 3C72B6ED22F3C614 CRC64;

Query Match
Best Local Similarity 40.5%; Score 190.5; DB 2; Length 996;
Matches 37; Conservative 15; Mismatches 32; Indels 1; Gaps 1;

QY 4 GVKVOYKNNDSAPGDNQIKPGQLVNTGSSVDLSVTYVRYWFTRDGSGSTLVYNCDAWAA 63
Db 844 GLKVLKNNETSASTGSIREFWFKVNGSSVDLSRVKIRYWTVDGDKPQSAV-CDWAAQ 902
QY 64 MCGCNIRASFGSVNPTATPTADTYLQ 88
Db 903 IGASNVTFNFVKLSGSGVADYYLE 927

RESULT 15
O52374 PRELIMINARY; PRT; 1779 AA.
AC O52374;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Family 10 xylanase (EC 3.2.1.8).
GN XVNC.
OS Caldicellulosiruptor sp. Rt69B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=70295;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rt69B.1;
RA Morris D.D., Gibbs M.D., Ford M., Thomas J., Bergquist P.L.;
RT "Family 10 and 11 xylanase genes from Caldicellulosiruptor sp. Rt69B.1.";
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
CC - SIMILARITY: BELONGS TO CELLULASE FAMILY F (FAMILY 10 OF GLYCOSYL HYDROLASES).
DR EMBL; AF036924; AAB95326.1; -.
DR PIR; T31085; T31085.
DR HSSP; Q06851; INBC.
```

Search completed: May 11, 2004, 12:09:28  
Job time : 12.385 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 12:04:16 ; Search time 4.3675 Seconds  
(without alignments)  
1052.023 Million cell updates/sec

Title: US-09-917-376-4

Perfect score: 470

Sequence: 1 VSGGVKQYKNDSPGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents\_AA.\*

1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.psp.\*

2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.psp.\*

3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.psp.\*

4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.psp.\*

5: /cgn2\_6/ptodata/2/iaa/PTCUS\_COMB.psp.\*

6: /cgn2\_6/ptodata/2/iaa/backfiles1.psp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	195.5	41.6	616	3	US-09-136-574A-47
2	194.5	41.4	1751	3	US-09-136-574A-44
3	191.5	40.7	1426	3	US-09-136-574A-43
4	172	36.6	700	2	US-07-862-588B-2
5	165.5	35.2	551	2	US-09-033-537A-1
6	165	35.1	167	5	PT-US95-13813-9
7	165	35.1	476	4	US-09-339-159B-4
8	165	35.1	493	3	US-09-198-956-10
9	165	35.1	493	3	US-09-198-985A-12
10	165	35.1	493	4	US-09-694-531-12
11	165	35.1	493	4	US-09-670-141-10
12	165	35.1	493	4	US-10-072-152-12
13	120.5	25.6	531	2	US-07-862-588B-7
14	114	24.3	162	1	US-08-048-164A-2
15	114	24.3	162	1	US-08-460-462-2
16	114	24.3	162	1	US-08-460-457-2
17	114	24.3	162	1	US-08-460-458-2
18	114	24.3	162	2	US-08-460-455-2
19	114	24.3	162	2	US-08-330-394A-2
20	114	24.3	163	3	US-09-006-636-7
21	114	24.3	163	3	US-09-006-632-7
22	114	24.3	163	4	US-09-325-274-7
23	113	24.0	382	3	US-09-277-716-22
24	113	24.0	382	4	US-09-609-161B-22
25	112	23.8	154	2	US-08-330-394A-29
26	112	23.8	156	2	US-08-330-394A-22
27	64	13.6	143	4	US-09-301-593-26

ALIGNMENTS

RESULT 1

US-09-136-574A-47

; Sequence 47, Application US/09136574A

; Patent No. 6294366

; GENERAL INFORMATION:

; APPLICANT: Farrington, Graham K.

; Anderson, Paige

; Gibbs, Moreland

; Bergquist, Peter

; Daniels, Roy

; Morgan, Hugh W.

; Williams, Diane P.

; TITLE OF INVENTION: Compositions and Methods for

; Treating Cellulose Containing

; Cellulase Enzyme Compositions

; NUMBER OF SEQUENCES: 49

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Howson and Howson

; STREET: Spring House Corporate Center, P.O. Box 457

; CITY: Spring House

; STATE: PA

; COUNTRY: USA

; ZIP: 19477

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/136,574A

; FILING DATE: 19-Aug-1998

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/932,571

; FILING DATE: September 19, 1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Bak, Mary E.

; REGISTRATION NUMBER: 31,215

; REFERENCE/DOCKET NUMBER: 1997US001/CIP

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 215-540-9200

; TELEFAX: 215-540-5818

; TELEX: <Unknown>

; INFORMATION FOR SEQ ID NO: 47:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 616 amino acids

; TYPE: amino acid

; TOPOLOGY: single

; MOLECULE TYPE: protein

; SEQUENCE DESCRIPTION: SEQ ID NO: 47:

Sequence 5, Appli  
Sequence 36, Appl  
Sequence 30, Appl  
Sequence 3, Appli  
Sequence 18, Appl  
Sequence 43, Appl  
Sequence 4640, Ap  
Sequence 2, Appli  
Sequence 38, Appl  
Sequence 33, Appl  
Sequence 32, Appl  
Sequence 8, Appli  
Sequence 19, Appl  
Sequence 2, Appli  
Sequence 26, Appl  
Sequence 68, Appl  
Sequence 68, Appl  
Sequence 68, Appl





```

US-07-862-588B-2
; Sequence 2, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linnaa
; APPLICANT: Schlein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5916796o No. 5916796disk of No. 5916796th America, Inc.
; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
; REFERENCE/DOCKET NUMBER: 3425.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 700 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-862-588B-2

Query Match          36.6%; Score 172; DB 2; Length 700;
Best Local Similarity 44.9%; Pred. No. 1.le-10;
Matches   40; Conservative 18; Mismatches 27; Indels    4; Gaps     3

QY      1 VSGGVKVQVYXND-SAPGDOIKPGLQLVNTGSSSDLSVTVVRYWTRDGGSTLVNC 607
Db       : : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
        549 VNSDLVVQVKGRNATDNQIKPHFNQNKGTSVPDLSSLTKLYFTKD--SAAANGWI 607

QY      60 DWAAAGCGNRASFGSVNPATPTADTYLQ 88
Db       ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
        608 DWA KLGSNIQISFGNHGA--DSDTYAE 634

RESULT 5
US-09-033-537A-1
; Sequence 1, Application US/09033537A
; Patent No. 5958083
; GENERAL INFORMATION:
; APPLICANT: Onishi, Masahiro
; APPLICANT: Fich, Merete
; APPLICANT: Toft, Annette Hanne
; APPLICANT: Sh lein, Martin
; TITLE OF INVENTION: Prevention Of Back-Staining
; TITLE OF INVENTION: In Stone Washing
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5958083o No. 5958083disk of No. 5958083th America, Inc.
```

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/13813
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: BAYER=3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 167 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; PCT-US95-13813-9

Query Match 35.1%; Score 165; DB 5; Length 167;
Best Local Similarity 36.1%; Pred. No. 1.1e-10;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKQVKNDSAPGDNQIKPGLQLVNTGSSVDLSVTYRVYFTFRDGGSTLVYVNC 60
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
5 VSGNLKVEFYNSPDDTNSINPQFKVTNTGSSAIDLSKLTLYYYTVDGQDQTFW-CD 63
QY 61 WAAM-----GCGNIRASFGSVNPATPTADTYIQ 88
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
64 HAAIIGNSYNGITSNVKGTFFVKSSSTNNADTYLE 100

RESULT 7
US-09-339-159B-4
; Sequence 4, Application US/09339159B
; Patent No. 6566114
; GENERAL INFORMATION:
; APPLICANT: Kauppinen, Markus
; APPLICANT: Schulein, Martin
; APPLICANT: Schnorr, Kirk
; APPLICANT: Andersen, Lene
; APPLICANT: Bjornvad, Mads
; TITLE OF INVENTION: No. 6566114el Mannanases
; FILE REFERENCE: 5440.204-US
; CURRENT APPLICATION NUMBER: US/09/339,159B
; CURRENT FILING DATE: 1999-06-24
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 4
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Bacillus
US-09-339-159B-4

Query Match 35.1%; Score 165; DB 4; Length 476;
Best Local Similarity 36.1%; Pred. No. 3.9e-10;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKQVKNDSAPGDNQIKPGLQLVNTGSSVDLSVTYRVYFTFRDGGSTLVYVNC 60
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
314 VSGNLKVEFYNSPDDTNSINPQFKVTNTGSSAIDLSKLTLYYYTVDGQDQTFW-CD 372
QY 61 WAAM-----GCGNIRASFGSVNPATPTADTYIQ 88
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
373 HAAIIGNSYNGITSNVKGTFFVKSSSTNNADTYLE 409

RESULT 8

```

```

US-09-198-956-10
; Sequence 10, Application US/09198956
; Patent No. 6165769
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels Erik K.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Schnorr, Kirk
; TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus
; FILE REFERENCE: 5377.200-US
; CURRENT APPLICATION NUMBER: US/09/198,956
; CURRENT FILING DATE: 1998-11-24
; EARLIER APPLICATION NUMBER: 1344/97
; EARLIER FILING DATE: 1997-11-24
; EARLIER APPLICATION NUMBER: 60/067,240
; EARLIER FILING DATE: 1997-12-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Bacillus licheniformis
US-09-198-956-10

Query Match 35.1%; Score 165; DB 3; Length 493;
Best Local Similarity 36.1%; Pred. No. 4.1e-10;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKQVKNDSAPGDNQIKPGLQLVNTGSSVDLSVTYRVYFTFRDGGSTLVYVNC 60
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
331 VSGNLKVEFYNSPDDTNSINPQFKVTNTGSSAIDLSKLTLYYYTVDGQDQTFW-CD 389
QY 61 WAAM-----GCGNIRASFGSVNPATPTADTYIQ 88
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
390 HAAIIGNSYNGITSNVKGTFFVKSSSTNNADTYLE 426

RESULT 9
US-09-198-955A-12
; Sequence 12, Application US/09198955A
; Patent No. 6187580
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. 6187580el Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/09/198,955A
; CURRENT FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 493

```

```
; TYPE: PRT
; ORGANISM: Clostridium thermocellum
US-09-198-955A-12

Query Match      35.1%; Score 165; DB 3; Length 493;
Best Local Similarity 36.1%; Pred. No. 4.1e-10;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKQVQKNDSPGDNQIKPGLQVLNTGSSVDLSTVTYRYWFTRDGGSSTLVYNC 60
Db 331 VSGNLKVEFYNSPDDTNSINPQFKVTNTGSSAIDLKLTLYRYYYTVDGQKDTFW-CD 389

QY 61 WAAM-----CGCNIRASFGSVNPATPTADTYLQ 88
Db 390 HAAIGSGSYNGITSNVKGTfVKMSSTNNADTYLE 426

RESULT 10
US-09-694-531-12
; Sequence 12, Application US/09694531
; Patent No. 6368843
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. 6368843el Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/09/694,531
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 09/198,955
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Clostridium thermocellum
US-09-694-531-12

Query Match      35.1%; Score 165; DB 4; Length 493;
Best Local Similarity 36.1%; Pred. No. 4.1e-10;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKQVQKNDSPGDNQIKPGLQVLNTGSSVDLSTVTYRYWFTRDGGSSTLVYNC 60
Db 331 VSGNLKVEFYNSPDDTNSINPQFKVTNTGSSAIDLKLTLYRYYYTVDGQKDTFW-CD 389

QY 61 WAAM-----CGCNIRASFGSVNPATPTADTYLQ 88
Db 390 HAAIGSGSYNGITSNVKGTfVKMSSTNNADTYLE 426

RESULT 11
US-09-670-141-10
; Sequence 10, Application US/09670141
```

```
; Patent No. 6429000
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels Erik K.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Schnorr, Kirk
; TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus
; FILE REFERENCE: 5377.200-US
; CURRENT APPLICATION NUMBER: US/09/670,141
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: 09/198,956
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Bacillus licheniformis
US-09-670-141-10

Query Match      35.1%; Score 165; DB 4; Length 493;
Best Local Similarity 36.1%; Pred. No. 4.1e-10;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKQVQKNDSPGDNQIKPGLQVLNTGSSVDLSTVTYRYWFTRDGGSSTLVYNC 60
Db 331 VSGNLKVEFYNSPDDTNSINPQFKVTNTGSSAIDLKLTLYRYYYTVDGQKDTFW-CD 389

QY 61 WAAM-----CGCNIRASFGSVNPATPTADTYLQ 88
Db 390 HAAIGSGSYNGITSNVKGTfVKMSSTNNADTYLE 426

RESULT 12
US-10-072-152-12
; Sequence 12, Application US/10072152
; Patent No. 6677147
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. 6677147el Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/10/072,152
; PRIOR FILING DATE: 2002-02-07
; PRIOR APPLICATION NUMBER: US/09/198,955
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 12
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Clostridium thermoecellum
US-10-072-152-12

Query Match      35.1%; Score 165; DB 4; Length 493;
Best Local Similarity 36.1%; Pred. No. 4.1e-10;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVQVQKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYVWFTTRDGGSTLVYVNC 60
Db 331 VSGNLKVFYNSPDSTNSINPQKVTWGTSSAIDLSKLTLYRYVYVVGQDKQTFW-CD 389

QY 61 WAAM-----GCCNIRASFGSVNPAPTADTYLQ 88
Db 390 HAAIGNSNGSYNGITSNVKGTFTVKSSSTNNADTYLE 426

RESULT 13
US-07-862-588B-7
; Sequence 7, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linnaa
; APPLICANT: Sch lein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5916796o No. 5916796disk of No. 5916796th America, Inc.
; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
; REFERENCE/DOCKET NUMBER: 3425.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 531 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-862-588B-7

Query Match      25.6%; Score 120.5; DB 2; Length 531;
Best Local Similarity 32.9%; Pred. No. 3.5e-05;
Matches 28; Conservative 18; Mismatches 36; Indels 3; Gaps 2;

QY 2 SGGVQVQKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYVWFTTRDGGSTLVYVNC 61
Db 432 TGNLVQVQVGTSDATDNQKPSFNKNGTTPVNLGSLKXXXXXKD-SPADMSCSIDW 490
```

```
QY 62 AAMGCGNIRASFGSVNPAPTADTY 86
; Sequence 2, Application US/08048164A
; Patent No. 5496934
; GENERAL INFORMATION:
; APPLICANT: Shoseyov, Oded
; APPLICANT: Shpiegl, Itai
; APPLICANT: Goldstein, Marc A.
; APPLICANT: Doi, Roy H.
; TITLE OF INVENTION: CELLULOSE BINDING DOMAIN
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/048,164A
; FILING DATE: 14-APR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7809-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 162 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-048-164A-2

Query Match      24.3%; Score 114; DB 1; Length 162;
Best Local Similarity 29.9%; Pred. No. 4.2e-05;
Matches 29; Conservative 22; Mismatches 34; Indels 12; Gaps 4;

QY 2 SGGVQVQKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYVWFTTRDGGSTLVYVNC 61
Db 3 TSSMVEFYNSKSAQNTSITPIKITNTSDNLNLDKVRYYTSDGTQQTFF-CDH 61

QY 62 AAMGCGN-----IRASP--GSVNPAPTADTYLQ 88
Db 62 AGALLGNSYVDNTSKVTANFVKETASP-TSTYDTYVE 97

RESULT 15
US-08-460-462-2
; Sequence 2, Application US/08460462
; Patent No. 5670623
; GENERAL INFORMATION:
; APPLICANT: Shoseyov, Oded
; APPLICANT: Shpiegl, Itai
; APPLICANT: Goldstein, Marc A.
; APPLICANT: Doi, Roy H.
; TITLE OF INVENTION: METHODS OF USE OF CELLULOSE BINDING DOMAIN PROTEINS
; NUMBER OF SEQUENCES: 21
```

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,462
FILING DATE: concurrently herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/048,164
FILING DATE: 14-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Mirock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7809-006
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 162 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-460-462-2

```

```

Query Match      24.3%; Score 114; DB 1; Length 162;
Best Local Similarity 29.9%; Pred. NO. 4.2e-05;
Matches 29; Conservative 22; Mismatches 34; Indels 12; Gaps 4;

QY      2 SGGVKVQYKNDSPGDNQIKGLQLVNTGSSVDLSVTYVYWFTRDGGSTLVYNCDW 61
DB      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
          3 TSSMSVEFYFNKSAQTSITPIIKITNTSDSLNDLVKRYYYTSDGTQGTFW-CDH 61
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      62 AAMGCGN-----IRASF--GSVNPATPTADTYLQ 88
DB      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
          62 AGALLGNSYDNTSKVTANFVKETASP-TSTYDTYVE 97
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Search completed: May 11, 2004, 12:11:21  
Job time : 5.3675 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 12:09:36 ; Search time 10.385 Seconds  
(without alignments)  
2378.773 Million cell updates/sec

Title: US-09-917-376-4

Perfect score: 470

Sequence: 1 VSGGVKQYKNNDSAPGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1140673 seqs, 277566755 residues

Total number of hits satisfying chosen parameters: 1140673

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US05\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	469	99.8	88	12	US-09-917-376-5
2	469	99.8	88	14	US-10-155-400-5
3	469	99.8	89	12	US-09-917-376-4
4	469	99.8	89	14	US-10-155-400-4
5	469	99.8	154	10	US-09-917-378-4
6	469	99.8	762	10	US-09-917-378-1
7	466	99.1	150	10	US-09-917-384-5
8	466	99.1	150	10	US-09-917-383-5
9	466	99.1	1043	10	US-09-917-384-6
10	466	99.1	1043	10	US-09-917-383-6
11	466	99.1	1228	10	US-09-917-384-1
12	466	99.1	1228	10	US-09-917-383-1
13	462	98.3	957	12	US-09-917-376-1
14	462	98.3	957	14	US-10-155-400-1
15	257	54.7	741	14	US-10-156-761-8100

16	190	40.4	508	15	US-10-369-493-23151	Sequence 23151, A
17	180	38.3	1621	14	US-10-185-990-10	Sequence 10, Appl
18	165	35.1	476	12	US-10-372-054-4	Sequence 4, Appl
19	165	35.1	493	12	US-10-655-433-12	Sequence 12, Appl
20	165	35.1	493	13	US-10-072-152-12	Sequence 12, Appl
21	165	35.1	599	10	US-09-955-555A-29	Sequence 29, Appl
22	155.5	33.1	1352	10	US-09-784-554B-2	Sequence 2, Appl
23	150.5	32.0	1350	10	US-09-784-554B-4	Sequence 4, Appl
24	130.5	27.8	1483	12	US-10-282-122A-51483	Sequence 51483, A
25	114	24.3	256	14	US-10-261-446-6	Sequence 6, Appl
26	113	24.0	382	10	US-09-808-898-22	Sequence 22, Appl
27	105.5	22.4	163	12	US-10-460-524-2	Sequence 2, Appl
28	78.5	16.7	1049	12	US-10-282-122A-49900	Sequence 49900, A
29	75.5	16.1	618	12	US-10-211-462-223	Sequence 223, App
30	75.5	16.1	986	9	US-09-747-835A-53	Sequence 53, Appl
31	75.5	16.1	986	12	US-10-312-313-53	Sequence 101, App
32	75.5	16.1	986	14	US-10-120-604-101	Sequence 101, App
33	75.5	16.1	986	14	US-10-225-567A-406	Sequence 406, App
34	75.5	16.1	986	16	US-10-398-458-3	Sequence 3, Appl
35	75.5	16.1	1131	16	US-10-398-458-2	Sequence 24, Appl
36	75.5	16.1	1325	9	US-09-747-835A-24	Sequence 24, Appl
37	75.5	16.1	1325	12	US-10-312-313-24	Sequence 20, Appl
38	75.5	16.1	1346	9	US-09-747-835A-20	Sequence 61, Appl
39	75.5	16.1	1346	9	US-09-747-835A-61	Sequence 25, Appl
40	75.5	16.1	1346	12	US-10-312-313-20	Sequence 25, Appl
41	75.5	16.1	1346	12	US-10-312-313-61	Sequence 25, Appl
42	75.5	16.1	1371	9	US-09-747-835A-25	Sequence 25, Appl
43	75.5	16.1	1371	12	US-10-312-313-25	Sequence 196885, A
44	75	15.4	161	12	US-10-424-599-196885	Sequence 377, App
45	72.5	15.4	69	9	US-09-764-860-377	

ALIGNMENTS

RESULT 1  
US-09-917-376-5  
; Sequence 5, Application US/09917376  
; Publication No. US20040038334A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOLYTICUS  
; FILE REFERENCE: 40197.4US01  
; CURRENT APPLICATION NUMBER: US/09/917,376  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 88  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
; FEATURE:  
; OTHER INFORMATION: Carbohydrate binding domain  
US-09-917-376-5

Query Match	99.8%	Score 469;	DB 12;	Length 88;
Best Local Similarity	100.0%	Pred. No. 1.5e-48;		
Matches	88;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
QY	1	VSGGVKQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSVTVRVWFTRDGSSLTLYNCD	60	
Db	1	VSGGVKQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSVTVRVWFTRDGSSLTLYNCD	60	
QY	61	WAAMCGNIRASFGSVNPATPTADTYLQ	88	
Db	61	WAAMCGNIRASFGSVNPATPTADTYLQ	88	
RESULT 2				

```
US-10-155-400-5
; Sequence 5, Application US/10155400
; Publication No. US20030108988A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: NREL 01-36A
; CURRENT APPLICATION NUMBER: US/10/155,400
; CURRENT FILING DATE: 2002-10-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 88
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; OTHER INFORMATION: Carbohydrate binding domain
US-10-155-400-5

Query Match          99.8%; Score 469; DB 14; Length 88;
Best Local Similarity 100.0%; Pred. No. 1.5e-48;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSLTVTVYWFTRDGGSSLLVYNC 60
Db 1 VSGGVKQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSLTVTVYWFTRDGGSSLLVYNC 60

QY 61 WAAMCGNIRASFSGSVNPATPTADTYLQ 88
Db 61 WAAMCGNIRASFSGSVNPATPTADTYLQ 88

RESULT 3
US-09-917-376-4
; Sequence 4, Application US/09917376
; Publication No. US2004003834A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: 40197.4US01
; CURRENT APPLICATION NUMBER: US/09/917,376
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 89
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (89)
; OTHER INFORMATION: Any amino acid
; OTHER INFORMATION: Carbohydrate binding domain
US-09-917-376-4

Query Match          99.8%; Score 469; DB 12; Length 89;
Best Local Similarity 100.0%; Pred. No. 1.5e-48;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSLTVTVYWFTRDGGSSLLVYNC 60
Db 1 VSGGVKQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSLTVTVYWFTRDGGSSLLVYNC 60

QY 61 WAAMCGNIRASFSGSVNPATPTADTYLQ 88
Db 61 WAAMCGNIRASFSGSVNPATPTADTYLQ 88

US-10-155-400-4
; Sequence 4, Application US/10155400
; Publication No. US20030108988A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: NREL 01-36A
; CURRENT APPLICATION NUMBER: US/10/155,400
; CURRENT FILING DATE: 2002-10-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 89
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (89)
; OTHER INFORMATION: Any amino acid
; OTHER INFORMATION: Carbohydrate binding domain
US-10-155-400-4

Query Match          99.8%; Score 469; DB 14; Length 89;
Best Local Similarity 100.0%; Pred. No. 1.5e-48;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSLTVTVYWFTRDGGSSLLVYNC 60
Db 1 VSGGVKQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSLTVTVYWFTRDGGSSLLVYNC 60

QY 61 WAAMCGNIRASFSGSVNPATPTADTYLQ 88
Db 61 WAAMCGNIRASFSGSVNPATPTADTYLQ 88

RESULT 5
US-09-917-378-4
; Sequence 4, Application US/09917378
; Publication No. US20030119093A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT MANNANASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: 40197.7US01
; CURRENT APPLICATION NUMBER: US/09/917,378
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 154
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Carbohydrate
; OTHER INFORMATION: binding domain
US-09-917-378-4

Query Match          99.8%; Score 469; DB 10; Length 154;
Best Local Similarity 100.0%; Pred. No. 2.9e-48;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy	61	WAAMCGCNIRASFGSVNPATPTADTYLQ	88
Dd	61	WAAMCGCNIRASFGSVNPATPTADTYLQ	88



```
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
US-09-917-384-6

Query Match          99.1%; Score 466; DB 10; Length 1043;
Best Local Similarity 98.9%; Pred. No. 6.8e-47;
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLYNCD 60
   |||:|||||
Db 477 VSGGLKVQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLYNCD 536
   |||:|||||

QY 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
   |||:|||||
Db 537 WAAMGCGNIRASFGSVNPATPTADTYLQ 564
   |||:|||||

RESULT 10
US-09-917-383-6
; Sequence 6, Application US/09917383
; Publication No. US20030104522A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULASE FROM ACIDOTHERMUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,383
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1043
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
US-09-917-383-6

Query Match          99.1%; Score 466; DB 10; Length 1043;
Best Local Similarity 98.9%; Pred. No. 6.8e-47;
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLYNCD 60
   |||:|||||
Db 477 VSGGLKVQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLYNCD 536
   |||:|||||

QY 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
   |||:|||||
Db 537 WAAMGCGNIRASFGSVNPATPTADTYLQ 564
   |||:|||||

RESULT 11
US-09-917-384-1
; Sequence 1, Application US/09917384
; Publication No. US20030096342A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULASE FROM ACIDOTHERMUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,384
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1228
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
US-09-917-383-1

Query Match          99.1%; Score 466; DB 10; Length 1228;
Best Local Similarity 98.9%; Pred. No. 8.3e-47;
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLYNCD 60
   |||:|||||
Db 584 VSGGLKVQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLYNCD 643
   |||:|||||

QY 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
   |||:|||||
Db 644 WAAMGCGNIRASFGSVNPATPTADTYLQ 671
   |||:|||||

RESULT 12
US-09-917-383-1
; Sequence 1, Application US/09917383
; Publication No. US20030104522A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULASE FROM ACIDOTHERMUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,383
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1228
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
US-09-917-383-1

Query Match          99.1%; Score 466; DB 10; Length 1228;
Best Local Similarity 98.9%; Pred. No. 8.3e-47;
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLYNCD 60
   |||:|||||
Db 584 VSGGLKVQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLYNCD 643
   |||:|||||

QY 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
   |||:|||||
Db 644 WAAMGCGNIRASFGSVNPATPTADTYLQ 671
   |||:|||||

RESULT 13
US-09-917-376-1
; Sequence 1, Application US/09917376
; Publication No. US20040038334A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS
; FILE REFERENCE: 40197.4US01
```

; CURRENT APPLICATION NUMBER: US/09/917,376  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 957  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
; FEATURE:  
; NAME/KEY: MOD RES  
; LOCATION: (957)  
; OTHER INFORMATION: Any amino acid  
US-09-917-376-1

Query Match 98.3%; Score 462; DB 12; Length 957;  
Best Local Similarity 97.7%; Pred. No. 1.9e-46;  
Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VSGGVKQVQYKXNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSSTLVYNC 60  
DB 869 VSGGVKQVQYKXNDSPAGDNQIKPGLQVNTGSSVDLSTVTYRYWFTRDGSSSTLVYNC 928  
QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
DB 929 WAAIGCGNIRASFGSVNPATPTADTYLQ 956

RESULT 14  
US-10-155-400-1  
; Sequence 1, Application US/10155400  
; Publication No. US2003010898A1  
; GENERAL INFORMATION:  
; APPLICANT: DING SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS  
; FILE REFERENCE: NREL 01-36A  
; CURRENT APPLICATION NUMBER: US/10/155,400  
; CURRENT FILING DATE: 2002-10-22  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 957  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
; FEATURE:  
; NAME/KEY: MOD RES  
; LOCATION: (957)  
; OTHER INFORMATION: Any amino acid  
US-10-155-400-1

Query Match 98.3%; Score 462; DB 14; Length 957;  
Best Local Similarity 97.7%; Pred. No. 1.9e-46;  
Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VSGGVKQVQYKXNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSSTLVYNC 60  
DB 869 VSGGVKQVQYKXNDSPAGDNQIKPGLQVNTGSSVDLSTVTYRYWFTRDGSSSTLVYNC 928  
QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
DB 929 WAAIGCGNIRASFGSVNPATPTADTYLQ 956

RESULT 15  
US-10-156-761-8100  
; Sequence 8100, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO

; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 8100  
; LENGTH: 741  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-8100

Query Match 54.7%; Score 257; DB 14; Length 741;  
Best Local Similarity 56.3%; Pred. No. 6e-22;  
Matches 49; Conservative 13; Mismatches 25; Indels 0; Gaps 0;  
QY 2 SGGVKVQYKXNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSSTLVYNC 61  
DB 591 SGGKLVLYKXNDSSATDNAIRPGLRVNTGSGSLDLKVTARYVFTRDGSGPTVNAWCDY 650  
QY 62 AAMGCGNIRASFGSVNPATPTADTYLQ 88  
DB 651 AAVGCSNVLKVVPLTTPVPGADAYLE 677

Search completed: May 11, 2004, 12:25:48  
Job time : 11.385 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 12, 2004, 03:40:05 ; Search time 1279.29 Seconds  
(without alignments)  
3015.368 Million cell updates/sec

Title: US-09-917-376-4

Perfect score: 470

Sequence: 1 VSGGVKQYKNDSPAGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frames\_p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US09917376/runat\_11052004\_114535\_28241/app\_query.fasta\_1.1429  
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09917376@cgn\_1\_7742@runat\_11052004\_114535\_28241 -NCPU=6 -ICPU=3  
-NO MAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOF=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.to.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.or.\*
- 21: em.ov.\*
- 22: em.pat.\*
- 23: em.ph.\*
- 24: em.pl.\*
- 25: em.ro.\*
- 26: em.sts.\*
- 27: em.un.\*
- 28: em.un.\*

RESULT 1

ALIGNMENTS

Result No.	Score	Query			DB ID	Description
		Match	Length	%		
1	469	99.8	2289	6	AX700050	Sequence
2	466	99.1	3687	6	AX700036	Sequence
3	462	98.3	2869	6	AX700058	Sequence
4	436	92.8	3365	6	AX700025	Sequence
5	257	54.7	299175	1	AP005023	Streptomy
6	228.5	48.6	1357	1	AY298814	Thermobif
7	225.5	48.0	4567	1	AF163837	Caldibaci
8	214.5	45.6	3237	1	AF200304	Caldibaci
9	209.5	44.6	3509	1	PAE488933	Paenibaci
10	204.5	43.5	4161	1	BSP133614	Bacillus
11	195.5	41.6	2029	6	E35142	Truncated c
12	195.5	41.6	2029	6	E35143	Truncated c
13	194.5	41.4	6005	1	AF078038S5	Caldicell
14	194.5	41.4	6416	6	E35100	Truncated c
15	192.5	41.0	4743	1	AB008029	Thermophi
16	192.5	41.0	11707	1	AF078737	Caldicell
17	192.5	41.0	11707	6	E35099	Truncated c
18	191.5	40.7	787	1	AF078038S3	Caldicell
19	191	40.6	5513	1	AT286105	A.thermophi
20	190.5	40.5	3262	1	AF078038S1	Caldicell
21	190.5	40.5	5437	1	CASR69XN2	Caldicell
22	190	40.4	1523	1	AF355629	Bacillus
23	190	40.4	1800	1	BSEGLSG	Bacillus su
24	190	40.4	1920	1	BACGLUB	B. subtilis
25	190	40.4	2314	1	BS14GLUC	B. subtilis
26	190	40.4	2435	1	BSBGLUC2	Bacillus su
27	190	40.4	2589	1	AY044252	Bacillus
28	190	40.4	26170	1	BC170DEGR	B. subtilis
29	190	40.4	207829	1	BSUB0010	Bacillus su
30	188	40.0	1928	1	BACCMCASE	D01057 B. subtilis
31	187	39.8	1593	1	AY183475	Bacillus
32	187	39.8	2084	1	AF045482	Bacillus
33	186.5	39.7	4977	1	CDCMANA	Caldocellum
34	186.5	39.7	5439	1	CDCELA	Caldocellum
35	186	39.6	2175	1	BACCELD	B. subtilis
36	186	39.6	2175	6	E05425	DNA sequenc
37	186	39.6	2587	1	AB016164	Bacillus
38	182.5	38.8	2977	6	A28170	B. lautus st
39	182.5	38.8	4241	1	CSCELB	Caldocellum
40	182.5	38.8	5284	1	CSU16308	Caldocellum
41	180.5	38.4	2831	1	BACCELA	Bacillus la
42	180	38.3	1553	1	BSU27084	Bacillus sp
43	177	37.7	2152	1	AF363635	Bacillus
44	172.5	36.7	6555	1	AF155197	Acetivibr
45	168	35.7	291000	1	SC0939105	Streptomy

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

```

AX700050
LOCUS AX700050 2289 bp DNA linear PAT 03-APR-2003
DEFINITION Sequence 2 from Patent WO03012110.
ACCESSION AX700050
VERSION AX700050.1 GI:29536020
KEYWORDS
SOURCE
ORGANISM
Acidothermus cellulolyticus
Acidothermus cellulolyticus
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Frankineae; Acidothermaceae; Acidothermus.
REFERENCE
1 Ding, S.Y., Adney, W.S., Vinzant, T.B. and Himmel, M.E.
AUTHORS Thermal tolerant mannanase from Acidothermus cellulolyticus
TITLE Patent: WO 03012110-A 2 13-FEB-2003;
JOURNAL Midwest Research Institute (US)
FEATURES
source
1..2289
/organism="Acidothermus cellulolyticus"
/mol_type="unassigned DNA"
/db_xref="taxon:28049"

ORIGIN
Alignment Scores:
Pred. No.: 2,218-48 Length: 2289
Score: 469.00 Matches: 88
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.79% Indels: 0
DB: 6 Gaps: 0

US-09-917-376-4 (1-89) x AX700050 (1-2289)

QY 1 ValSerGlyValLysValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGln 20
Db 1363 GTGTCGGGTGGGTGAAGTGCAGTACAGACATGATTCGGCGCGGTGATAACCA 1422
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1423 ATCAACCGGGTCTCCAGTTGGTGAATACGGGGTCTGTCGGTGGATTTCGACGG 1482
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1483 ACGGTGGCGGTACTGGTTACCCGGATGTGGTCTGCACACTGGTGTAACACTGTGAC 1542
QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 1543 TGGGGCGGCGATGGGTGTGGATATCCGGCTCGTTCGGTTCGGTGAACCCGGCGAC 1602
QY 81 ProThrAlaAspThrTyrLeuGln 88
Db 1603 CCGACGGCGGACACCTACCTGCAG 1626

RESULT 2
AX700036
LOCUS AX700036 3687 bp DNA linear PAT 03-APR-2003
DEFINITION Sequence 2 from Patent WO03012109.
ACCESSION AX700036
VERSION AX700036.1 GI:29536019
KEYWORDS
SOURCE
ORGANISM
synthetic construct
synthetic construct
artificial sequences.
REFERENCE
1 Ding, S.Y., Adney, W.S., Vinzant, T.B., Himmel, M.E. and Decker, S.R.
AUTHORS Thermal tolerant cellulase from Acidothermus cellulolyticus
TITLE Patent: WO 03012109-A 2 13-FEB-2003;
JOURNAL Midwest Research Institute (US)
FEATURES
source
1..3687
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="Segment of Guxa"

```

```

ORIGIN
Alignment Scores:
Pred. No.: 9,538-48 Length: 3687
Score: 466.00 Matches: 87
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 98.86% Mismatches: 0
Query Match: 99.15% Indels: 0
DB: 6 Gaps: 0

US-09-917-376-4 (1-89) x AX700036 (1-3687)

QY 1 ValSerGlyValLysValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGln 20
Db 1750 GTGTCGGGTGGGTGAAGTGCAGTACAGACATGATTCGGCGCGGTGATAACCA 1809
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1810 ATCAACCGGGTCTCCAGTTGGTGAATACCGGGTCTGTCGGTGGATTTCGACCG 1869
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1870 ACGGTGGCGGTACTGGTTACCCGGATGTGGTCTGCACACTGGTGTAACACTGTGAC 1929
QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 1930 TGGGGCGGCGATGGGTGTGGATATCCGGCTCGTTCGGTTCGGTGAACCCGGCGAC 1989
QY 81 ProThrAlaAspThrTyrLeuGln 88
Db 1990 CCGACGGCGGACACCTACCTGCAG 2013

RESULT 3
AX700058
LOCUS AX700058 2869 bp DNA linear PAT 03-APR-2003
DEFINITION Sequence 2 from Patent WO03012090.
ACCESSION AX700058
VERSION AX700058.1 GI:29536021
KEYWORDS
SOURCE
ORGANISM
Acidothermus cellulolyticus
Acidothermus cellulolyticus
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Frankineae; Acidothermaceae; Acidothermus.
REFERENCE
1 Ding, S.Y., Adney, W.S., Vinzant, T.B. and Himmel, M.E.
AUTHORS Thermal tolerant avicelase from Acidothermus cellulolyticus
TITLE Patent: WO 03012090-A 2 13-FEB-2003;
JOURNAL Midwest Research Institute (US)
FEATURES
source
1..2869
/organism="Acidothermus cellulolyticus"
/mol_type="unassigned DNA"
/db_xref="taxon:28049"

ORIGIN
Alignment Scores:
Pred. No.: 2,198-47 Length: 2869
Score: 462.00 Matches: 86
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 97.73% Mismatches: 0
Query Match: 98.30% Indels: 0
DB: 6 Gaps: 0

US-09-917-376-4 (1-89) x AX700058 (1-2869)

QY 1 ValSerGlyValLysValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGln 20
Db 2605 GTGTCGGGTGGGTGAAGTGCAGTATAGAAATATGATTCGGCGCGGTGATAACCA 2664
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 2665 ATCAACCGGGTTCGACGTGGTGAATACCGGGTCTGTCGGTGGATTTCGACCGGTG 2724

```

```

Qy 41 ThrValArgTyrTrpPheThrArgAspGlySerSerThrLeuValTyrAsnCysAsp 60
Db 2725 ACGTTCGGTACTGCTTCAACCGGGATGGTGGCTCGTCACACTGGTGTAACAATGTGAC 2784

Qy 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 2785 TGGGCGGGAGTACGGGTGTGGGATATCCGCCCTCGTTCCGCTCGGTGAACCCGGCGAG 2844

Qy 81 ProThrAlaAspThrTyrLeuGln 88
Db 2845 CCGACGGGGACACCTACCTGCAG 2868

RESULT 4
AX700025 3365 bp DNA linear PAT 03-APR-2003
LOCUS Sequence 2 from Patent WO03012095.
DEFINITION AX700025
ACCESSION AX700025
VERSION AX700025.1 GI:29536018
KEYWORDS
SOURCE
ORGANISM
Acidothermus cellulolyticus
Bacteria: Actinobacteria: Actinobacteridae; Actinomycetales;
Frankineae; Acidothermaceae; Acidothermus.
REFERENCE
1 Adney,W.S., Ding,S.Y., Vinzant,T.B., Himmel,M.E., Decker,S.R. and
Lantz McCarter,S.
Thermal tolerant exoglucanase from Acidothermus cellulolyticus
Patent: WO 03012095-A 2 13-FEB-2003;
Midwest Research Institute (US)
LOCATION/Qualifiers
1. .3365
/organism="Acidothermus cellulolyticus"
/mol_type="unassigned DNA"
/db_xref="taxon:28049"

ORIGIN
Alignment Scores:
Pred. No.: 4.72e-44 Length: 3365
Score: 436.00 Matches: 81
Percent Similarity: 97.62% Conservative: 1
Best Local Similarity: 96.43% Mismatches: 2
Query Match: 92.77% Indels: 0
DB: 6 Gaps: 0

US-09-917-376-4 (1-89) x AX700025 (1-3365)

Qy 5 VallysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysProGly 24
Db 112 CTCAAAGCGCAGTATAAGACAATGATTCGGCGCGGAGTGACACACGATCAACCGGGT 171

Qy 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArgTyr 44
Db 172 CTCACAGTGGTGAATACCGGTGCTCGCGTGGATTGTTCGACGGTACCGTGGCGTAC 231

Qy 45 TrpPheThrArgAspGlySerSerThrLeuValTyrAsnCysAspTrpAlaAlaMet 64
Db 232 TGGTTACCCCGGGATGGTGGGTCTGTCGACACTGGTGTAACATGTAATGGCGCGGATG 291

Qy 65 GlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAlaAsp 84
Db 292 GGGTGTGGGAATATCCGGCCTCGTTCCGCTCGGTGAACCCGGCGACCGCGCGGAC 351

Qy 85 ThrTyrLeuGln 88
Db 352 ACCTACCTGCAG 363

RESULT 5
AP005023 299175 bp DNA linear BCT 10-MAY-2003
LOCUS Streptomyces avermitilis genomic DNA, complete genome, section
DEFINITION AP005023
ACCESSION AP005023 BA000030
VERSION AP005023.1 GI:29604083

```

## KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

Streptomyces avermitilis MA-4680  
Streptomyces avermitilis MA-4680  
Bacteria: Actinobacteriae; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomycetes.

1 Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T., Kikuchi, H., Shiba, T., Sakaki, Y. and Hattori, M.  
Genome sequence of an industrial microorganism Streptomyces avermitilis: deducing the ability of producing secondary metabolites  
Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)  
21477403  
11572948

2 Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M. and Omura, S.  
Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermitilis  
Nat. Biotechnol. 21 (5), 526-531 (2003)  
22608306  
12692562

3 (bases 1 to 299175)  
Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M., Director-General of Biotechnology Center, Shiba, T., Sakaki, Y. and Hattori, M.  
Direct Submission  
Submitted (29-MAR-2002) Director-General of Biotechnology Center, National Institute of Technology and Evaluation, Biotechnology Center; 2-49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan  
(E-mail: bio@nitech.go.jp, URL: http://www.bio.nitech.go.jp/, Tel. 81-3-3481-1933, Fax: 81-3-3481-8424)  
This work was done in collaboration with Haruo Ikeda(\*1), Jun Ishikawa(\*2), Akihito Hanamoto(\*3), Chigusa Takahashi(\*3), Mayumi Shinose(\*3), Hiroshi Horikawa(\*4), Hidekazu Nakazawa(\*4), Tomomi Osonoe(\*4), Norihiro Kishida(\*4), Hisashi Kikuchi(\*4), Tadayoshi Shiba(\*5), Yoshiyuki Sakaki(\*6,\*7), Masahira Hattori(\*1,\*7) and Satoshi Omura(\*1,\*3).  
Final finishing process and all annotation were done by H. Ikeda and J. Ishikawa.  
\*1 Kitasato Institute for Life Sciences, Kitasato University  
\*2 National Institute of Infectious Diseases  
\*3 The Kitasato Institute  
\*4 National Institute of Technology and Evaluation  
\*5 School of Science, Kitasato University  
\*6 Institute of Medical Science, University of Tokyo  
\*7 RIKEN, Genomic Sciences Center  
Following url is also available.  
http://avermitilis.ls.kitasato-u.ac.jp.  
Location/Qualifiers  
1. .299175  
/organism="Streptomyces avermitilis MA-4680"  
/mol\_type="genomic DNA"  
/strain="MA-4680"  
/db\_xref="taxon:227882"  
/note="This strain is also named as strain: ATCC 31267, NCIMB 12804 or NRRL 8165.-synonym: Streptomyces avermectinus"  
complement(155..550)  
/note="SAV445"  
Tn4561, truncated"  
/codon\_start=1  
/transl\_table=11  
/product="putative transposase"  
/protein\_id="BAC68155.1"  
/db\_xref="GI:29604084"  
/translation="MAGPPRSATRSRTTADREDSAPADRADEPGYRRQIKVQANLQ EGHVARRAGQLYQRYQDMEDQIGALGLVNLVLFNTRYMDALTQLRADGFEVR DDHVARLSPFMRGHYINVLGRYSFQLPDL"  
complement(474..740)  
/note="IS630 family, truncated"

```
SAV446"
/codon_start=1
/transl_table=11
/product="putative transposase"
/protein_id="BAC68156.1"
/db_xref="GI:29604085"
/translation="NPDTQAGPSTVTAATAVSTAGSYVSMAPTSFVWAPPSGRRR
EPGLGAVRAYDVIRMLSDGRPTPLGDAITYGGSRLCTSCA"
766..987
/note="SAV447"
/codon_start=1
/transl_table=11
/product="putative transposase"
/protein_id="BAC68157.1"
/db_xref="GI:29604086"
/translation="MVVAVGCAVSTVPTAGGVAAEPCVVRRLTDQEGKLOQIVRRG
SASSVRRFRAMMLASAGNRPVIAQLV"
complement(1954..2571)
/note="SAV448"
/codon_start=1
/transl_table=11
/product="putative TetR-family transcriptional regulator"
/protein_id="BAC68158.1"
/db_xref="GI:29604087"
/translation="MPTSTRARERANTRERIEAALHVLTEGTAALTIRIATDVE
YAPVYQHFANKDALVLELVAHGRMLSEFSQAQPDPTDRRMTIASQYRFAGE
HPHIFQVNDPMDVADERRVAEPVIDVLEKLLTWSAHDVVLADFQACEILWGL
YGIASLSYLGSVGNDRLARYLAEQALHTILLGWRTPANGRPASS"
2694..3098
/note="SAV449"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC68159.1"
/db_xref="GI:29604088"
/translation="MIYHNRVTMKATPTEQIEGVLESRWNRQSNPAIKSFVVGRN
HGDTYGAVFVEHLGLFAYLTHPTTYQTDHLHLVERLEIFDVSDDDPDNLAK
IQELHRRNELNPQIAGMLADVPTTYTGSVD"
3127..4029
/note="SAV450"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC68160.1"
/db_xref="GI:29604089"
/translation="NAKXVIIGAGSIGTNAFLAEGGESVRIIVTRSGSGPEHPLVDR
VAADSPSRITELSRGAVIYHCANPSYTWERLLPLPQTAAITAKANDAVLALT
GSLYAYGPQGRNMEHTPMAATGHKRLRRMWEQALAGGITRVEVRSYIGKDAN
GIYSLIIKSLAQKGAWIITGHDMPHTFTFNGDMAQALITLASERAWGRPWVPSA
PAVITRELARYAVAAGRPPIKLIQIPRSVTRTAGLIVPIAREMAEMDYQWYAPFMD
ATETTFGLTATDLDTAVRNEVG"
complement(4130..4498)
/gene="bldg"
/complement(4130..4498)
/gene="bldg"
/note="SAV451"
/codon_start=1
/transl_table=11
/product="putative anti-anti-sigma factor"
/protein_id="BAC68161.1"
/db_xref="GI:29604090"
/translation="WTDAPFIDVPHTDAAVIALSGRFDIAAAPAVRARALELIANG
HPDLVADISGVTFCDSSGLGALVGIWRCADGSLTLAAIPDLRSLRLSVTGMDAFL
PAYASDAALAAQRNRITA"
4635..5156
/note="SAV452"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC68162.1"
/db_xref="GI:29604091"
/translation="MFVAKESRVVVGIAATPTGEALELVLHLGESLRGVRESDDA
TARPDGLDLAELPARGEWHTTALVVRPSVEGEAGGLFHPGAEPGESGPARRLCAWRE
```

```
SAPRTDIEDPEVHRTLRRSLEDLGAPAHLRPAAPRTHGGRRRCPLVHGVPVRPLAAV
LVHGTAAAGPCTPR"
5183..5665
/note="SAV453"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC68163.1"
/db_xref="GI:29604092"
/translation="MVTAVRARAGSAMQACRGARRSSQHSSEGVRRMDEKEFLRTV
ABRTGLSROEAADLTWATLETLARRLSPGEARDLITELPGLAEAVRGTTDRTERFG
HSDSVRRVAERTMLKEEADRGVTVLAVLREALSEKFNDLSQLGTDTSQALLESAG
"
5810..6964
/note="SAV454"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC68164.1"
/db_xref="GI:29604093"
/translation="MIMTADVPETSSGRSVRVVALVADPGLPSEIARALAPRLPDHLRR
AVCARVMQVDTMTAPLSVSEQADVSDIAEAVRSHLDGAEDVAIFUTDPRRARLYP
ISVEVDTALRAALISLPALGVRLRRVRQAVDVVRELVAHEETPSHPVEVIGRPPE
PUDTRPGEQQRVYVVGRLGHTRLVAGMYANRPGLEGSRLAGLAPATATLVF
INSATWNLATAGAVQHVATTAGSAVALTLWIIIVDHLWERTGLHPARHHPILYTNLV
TLVTTALAVVLAALFATLVAVSFLLDASVLKFKTGRQVGVGDYLFLLAWFITAIAM
IGAFGTSLEGDEKVSDAAYGRQRQRMLRQMPRPERGERRPGDPTTDG"
complement(6986..7498)
/note="SAV455"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC68165.1"
/db_xref="GI:29604094"
/translation="MNPYPASGIEQRPQHSWLDLYWCSCDIIADLAEGLEORPLV
ATVAREAGERVLAQVPAQRVTWRALGDGRYHQNVAFGSPAVFLVMSMASMSALG
NARRRAAADAQPRCCARTPGLHVLVDLVHGTMLRSMHLGTSSTRPSSAURLAL
ELVDLHVIAL"
7548..8189
/note="SAV456"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC68166.1"
/db_xref="GI:29604095"
/translation="MATLRRVCPHEGVRTAAVPAQFVLSRRPWTASVYRCVAVGPOS
ERRVEGSRTPVASASSRLRPCRRLSVVRCPRAVPSQAVRSRFRGTRTFLGDVL
VYORHDLAHLVRHVDDVHPGLGRGAAPVRIATHSLGGLVIAVDWATANVPLWES
LVTFGSQATFFHQVPEGDLQCQIAFPMRKHQAETGALPPETGFIHAGQGW"

Alignment Scores:
Pred. No.:      2,91e-19      Length:      299175
Score:          257.00        Matches:      49
Percent Similarity: 71.26%    Conservative: 13
Best Local Similarity: 56.32%  Mismatches:   25
Query Match:     54.68%      Indels:       0
DB:              1           Gaps:         0

US-09-917-376-4 (1-89) x AP005023 (1-299175)

QY      2  SerGlyGlyValIysValGIntYrLysAsnAsnAspSerAlaProGlyAspAsnGlnIle 21
Db      105341 TCGGGGGGCTCAAGGTCCTCTACAGAACACACGACTCTCTCGGCCACCGACACGCCATC 105400
QY      22  LysProGlyLeuGlnLeuValAsnThrGlySerSerSerValAspLeuSerThrValThr 41
Db      105401 CGGCCAGGCGCTTCGAGTCGTCAACACCGCGAGCGGCTCCCTCGACCTGTCCAGGTCACG 105460
QY      42  ValArgTy-TrpPheThrArgAspGlyGlySerSerThrLeuValTyAsnCysAspTrp 61
Db      105461 GCCCGGTACTACTACTACCCGGGACAGCGGCTCGCCCAACCGCTGTGGTGGGACTAC 105520
QY      62  AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81
```

DB: 1 2 Gaps:

US-09-917-376-4 (1-89) x AY298814 (1-1957)

QY 2 SerGlyValLeuValGlnTyrLysAsnAspSerAlaProGlyAspAenGlnIle 21  
Db :  
1517 ACGGGTCCTCAGGTCTACTACCCCAACAACAGTCTTGGGCCGCACGACCCAGATC 1576

QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41  
Db :  
1577 GCGCGGGCTCGTCTGGTCAACACCGAGCAGCACGGTCGACCTGGCTGACGTGGAA 1636

QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTyr 61  
Db :  
1637 ATCCACTACTACTTACCACGAGCCCGGC--GGTACCTCCAGTTCACTCCGACTGG 1693

QY 62 AlaAlaMetGlyCysglyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81  
Db :  
1694 GCTCAAGTGGGTGCGCCCAACGCTCAACGGGTCTTCACGTGCTG-----TCGGTCCG 1747

QY 82 ThrAlaAspThrTyrLeu 87  
Db :  
1748 GGCGCCGACACTCCCTG 1765

RESULT 7  
AF163837 4567 bp DNA linear BCT 08-FEB-2000  
LOCUS  
DEFINITION  
Calidibacillus cellulovorans multidomain beta-1,4-mannanase  
precursor (manA) gene, complete cds; and unknown genes.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AF163837.1 GI:6651325  
. Calidibacillus cellulovorans  
Calidibacillus cellulovorans  
Bacteria; Firmicutes; Bacillales; Alicyclobacillaceae;  
Calidibacillus.  
1 (bases 1 to 4567)  
Sunna,A., Gibbs,M.D., Chin,C.W., Nelson,P.J. and Bergquist,P.L.  
A gene encoding a novel multidomain beta-1,4-mannanase from  
Calidibacillus cellulovorans and action of the recombinant enzyme on  
kraft pulp  
Appl. Environ. Microbiol. 66 (2), 664-670 (2000)

JOURNAL MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES

source  
1. 4567  
/organism="Calidibacillus cellulovorans"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:74586"  
<1..513  
/notes="orf1"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAF22273.1"  
/gb\_xref="GI:6651326"  
/translation="TPPTPTPSAAPTPTPSAGSLVVOYRAADTNAGDNOLKPHPIR  
VNRGSVPSELSTIRYWTVDGDKPOVCNDAWGCVCSNVGRSFVKLSLTGRGTGDY  
ISITFSGAGSLAAGASSDGIQRINKNDWNTNEANDYSIDPTKTSFADNRRTLYLR  
NGQLIWGVSP"  
1..60  
/notes="Region: proline-threonine linker"  
/evidence="not\_experimental"

misc\_feature  
61..510  
/notes="Region: cellulose-binding domain type IIIB"  
/evidence="not\_experimental"

gene  
620..3412  
/genes="manA"

CDS	620..3412 /gene="mana" /codon_start=1 /transl_table=11 /product="multidomain beta-1,4-mannanase precursor" /protein_id="AAF22274.1" /db_xref="GI:6651327"	DB:	1	Gaps:	1
	/translation="MNRRLIARLSGLMLAWLIAVLAVPKPEPAEAGHGMVFPATRT YACYYDGVHNGGDLNMINPACLDALAIISGNYQFWMFGLNISNAGGRHREIIPDGK LCGTPADGNDQARTDMWTRRLQPGATITVRVNAWAPHGPTWYLYVRDGDWTPQPL KWSLEPPTFFVONPPINSSGPDGAESYQVQLPNKQGRHIIYIMWORSDFEAFYN CSDVYEGSPPIAYERGDREGGWTTPPSGTTPTPTPTPTPTPTPTPTPTPTPTPTPT TPTSTPTPSAGSTLAEYVRVGDTSATDNQMPQLAIVNTGSOAVLELTKVRYWTKN STQAEYFCDDWAQGCNSIRAQFVLSQFVSGADSYIELSFTGSGIPAGNTEIQNR IHFTNMNNTDMSYNAQTWGPSTRIILYRNGVLVWGTEPGGSSPPTPTPTPT GANAARIIVLSNGCRWTKIPASEVADIISOARTGLYRAVLEVHDFTGVEDAAACSMT TAVNYIELKVLVLAQENFIVNIGNEPYGNVNYQWVTDFTENAVOALRNAGINNTIM VDAPNWGDWSTWRDNAPTFINADPQRLNLFVSIHNYGYDTAAEVQSYIESFVNRGL PLVIGFEGHMSDFDNEQAIQVYAKQINIGLFGWSGNGGVGEYLDVMTNFNANS TAWGTWPTNAIGTSTPT KPHFRIVNRGTSSVPLSELTRYWTVYVDGDKPQVFNCDWAQVGCNLRGSGFKLSTGR TGADYIIBITFTSGAGSLAPGASSGDIQVRINKNDWTNNEANDYSYDPTKTSFADWN RVTLRYNGQLVWGVEP"	US-09-917-376-4 (1-89) x AF163837 (1-4567)			
sig_peptide	620..718 /gene="mana" /evidence=not_experimental	QY	3	GlyGlyValIysValGlnTyrIysAsnAsnAspSerValaProGlyAspAsnGlnIleIys	22
mat_peptide	719..3409 /gene="mana" /product="multidomain beta-1,4-mannanase" /gene="mana" /evidence="Region: proline-threonine linker"	Db	58	GGCAGCTGGTGTGAGTATCGCGCGGACACCAACGCGGCGACACACCTGAAG	117
misc_feature	1349..1480 /evidence="not_experimental"	QY	23	ProGlyLeuGlnLeuValAsnThrGlySerSerValaAspLeuSerThrValThrVal	42
misc_feature	1481..1927 /evidence="not_experimental"	Db	118	CCGATTTTTCGATTGTGAACCGCGGACGTCGAGCGTGCCTGTCGGAGCTCACGATT	177
misc_feature	1928..2032 /evidence="not_experimental"	QY	43	ArgTyrTyrPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTyrAla	62
misc_feature	2033..2884 /evidence="not_experimental"	Db	178	CGGTACTGTACTGACGGTGGAC---GGGGACAAGCCGAGGTGTCACTGCGACTGGCG	234
misc_feature	2885..2959 /evidence="not_experimental"	QY	63	AlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValaAsnProAlaThrProThr	82
misc_feature	2960..3409 /evidence="not_experimental"	Db	235	CAGTGGGTGTTTCGACGTGCGCGGACGTTCTGTAAGCTTTCGACGGGCGGACGGGG	294
		QY	83	AlaAspThrTyrIleuGln	88
		Db	295	GCGGACTACTATATCGAG	312
		RESULT 8			
		AF200304			
		LOCUS			
		DEFINITION			
		ACCESSION			
		VERSION			
		KEYWORDS			
		SOURCE			
		ORGANISM			
		REFERENCE			
		AUTHORS			
		TITLE			
		JOURNAL			
		MEDLINE			
		PUBMED			
		REFERENCE			
		AUTHORS			
		TITLE			
		JOURNAL			
		FEATURES			
		source			
		CDS			
		gene			
		CDS			
		Alignment Scores:			
		Pred. No.:			
		Score:			
		Percent Similarity:			
		Best Local Similarity:			
		Query Match:			







AUTHORS Paiji,A., Petaer,L.B., Roy,M.D., Gurahamu,K.F., Moreland,D.G.,  
Hyu,M. and Daian,P.W.  
TITLE Truncated cellulase composition  
JOURNAL Patent: JP 199221086-A 44 17-AUG-1999;  
CLARIANT INTERNATIONAL LTD  
COMMENT OS Artificial Sequence  
PN JP 199221086-A/44  
PD 17-AUG-1999  
PF 21-SEP-1998 JP 199221086-A/44  
PR 19-SEP-1997 US 08/932571  
PI PAIJI ANDERSON,PETAER L. BAGUKUISUTO,ROY M DANIEL, PI  
GURAHAMU K PARINTON,  
PI MORELAND DAVID GIBUSU,HYU MORGAN,DAIAN PURATONOTISU WILLIAM  
PC C12N15/09,C11D3/386,C12N1/21,C12N9/42/(C12N1/21,C12R1:19), PC  
(C12N9/42,C12R1:19),C12N15/00  
CC Location/Qualifiers  
FH Key 1..2029  
FT source /organism='Artificial Sequence'.  
FT Location/Qualifiers  
FEATURES  
source  
1..2029  
/organism='synthetic construct'  
/mol\_type='genomic DNA'  
/db\_xref='taxon:32630'  
ORIGIN  
Alignment Scores:  
Pred. No.: 2,6e-14 Length: 2029  
Score: 195.50 Matches: 38  
Percent Similarity: 60.23% Conservative: 15  
Best Local Similarity: 43.18% Mismatches: 34  
Query Match: 41.60% Indels: 1  
DB: 6 Gaps: 1  
US-09-917-376-4 (1-89) x E35142 (1-2029)  
QY 1 ValSerGlyGlyValIynValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20  
Db 1 ATGGGAAGTGGTGTGAAGTACTGTACAAAGCAATGAGCAAGTGGCGAGCACAGTTCT 60  
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40  
Db 61 ATAAGCCGCTGTTTAAGATAGTGAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 120  
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
Db 121 AAGATAAGATATCTGGTACACAGTGGATGGTGAACAGCCACAGAGTGGCGGTA---TGTGAC 177  
QY 61 TtpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80  
Db 178 TGGGCACAGATAGGGGCAAGCAATGTGACATTCATTTGTGAAGCTTAGCAGCGGAGTG 237  
QY 81 ProThrAlaAspThrTyrLeuGln 88  
Db 238 AGTGGAGCGGATTATTACCTGGAG 261  
RESULT 12  
E35143 2029 bp DNA linear PAT 18-JUN-2001  
LOCUS Truncated cellulase composition.  
DEFINITION E35143  
ACCESSION E35143  
VERSION E35143 1 GI:13018968  
KEYWORDS JP 199221086-A/45.  
SOURCE unclassified  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 2029)  
AUTHORS Paiji,A., Petaer,L.B., Roy,M.D., Gurahamu,K.F., Moreland,D.G.,  
Hyu,M. and Daian,P.W.  
TITLE Truncated cellulase composition  
JOURNAL Patent: JP 199221086-A 45 17-AUG-1999;  
CLARIANT INTERNATIONAL LTD  
COMMENT OS Unidentified

PN JP 199221086-A/45  
PD 17-AUG-1999  
PF 21-SEP-1998 JP 199221086-A/45  
PR 19-SEP-1997 US 08/932571  
PI PAIJI ANDERSON,PETAER L. BAGUKUISUTO,ROY M DANIEL, PI  
GURAHAMU K PARINTON,  
PI MORELAND DAVID GIBUSU,HYU MORGAN,DAIAN PURATONOTISU WILLIAM  
PC C12N15/09,C11D3/386,C12N1/21,C12N9/42/(C12N1/21,C12R1:19), PC  
(C12N9/42,C12R1:19),C12N15/00  
CC Location/Qualifiers  
FH Key 1..2029  
FT source /organism='Unidentified'.  
FT Location/Qualifiers  
FEATURES  
source  
1..2029  
/organism='unidentified'  
/mol\_type='genomic DNA'  
/db\_xref='taxon:32644'  
ORIGIN  
Alignment Scores:  
Pred. No.: 2,6e-14 Length: 2029  
Score: 195.50 Matches: 38  
Percent Similarity: 60.23% Conservative: 15  
Best Local Similarity: 43.18% Mismatches: 34  
Query Match: 41.60% Indels: 1  
DB: 6 Gaps: 1  
US-09-917-376-4 (1-89) x E35143 (1-2029)  
QY 1 ValSerGlyGlyValIynValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20  
Db 1 ATGGGAAGTGGTGTGAAGTACTGTACAAAGCAATGAGCAAGTGGCGAGCACAGTTCT 60  
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40  
Db 61 ATAAGCCGCTGTTTAAGATAGTGAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 120  
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
Db 121 AAGATAAGATATCTGGTACACAGTGGATGGTGAACAGCCACAGAGTGGCGGTA---TGTGAC 177  
QY 61 TtpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80  
Db 178 TGGGCACAGATAGGGGCAAGCAATGTGACATTCATTTGTGAAGCTTAGCAGCGGAGTG 237  
QY 81 ProThrAlaAspThrTyrLeuGln 88  
Db 238 AGTGGAGCGGATTATTACCTGGAG 261  
RESULT 13  
AF078038S 6005 bp DNA linear BCT 11-FEB-2001  
LOCUS Caldicellulosiruptor sp. Tok7B.1 glycosyl hydrolase 6 gene, partial  
DEFINITION cds; and CelE gene, complete cds.  
ACCESSION AF078042  
VERSION AF078042.1 GI:12743878  
KEYWORDS 5 of 5  
SEGMENT Caldicellulosiruptor sp. Tok7B.1  
SOURCE Caldicellulosiruptor sp. Tok7B.1  
ORGANISM Bacteria; Firmicutes; Clostridia; Clostridiales;  
Syntrophomonadaceae; Caldicellulosiruptor.  
REFERENCE 1 (bases 1 to 6005)  
AUTHORS Gibbs,M.D., Reeves,R.A., Farrington,G.K., Anderson,P.,  
Williams,D.P. and Bergquist,P.L.  
TITLE Multidomain and multifunctional glycosyl hydrolases from the  
extreme thermophile Caldicellulosiruptor isolate Tok7B.1  
JOURNAL Curr. Microbiol. 40 (5), 333-340 (2000)  
MEDLINE 20171169  
PUBMED 10706665  
REFERENCE 2 (bases 1 to 6005)  
AUTHORS Gibbs,M.D., Reeves,R.A., Farrington,G.K., Anderson,P.,

misc_feature		/note="PT-linker"	
3107..3223		/note="PT-linker"	
misc_feature		4196..4295	
/note="PT-linker"			
ORIGIN			
Alignment Scores:			
Pred. No.:	1.36e-13	Length:	6005
Score:	194.50	Matches:	38
Percent Similarity:	61.1%	Conservative:	14
Best Local Similarity:	44.71%	Mismatches:	32
Query Match:	41.38%	Indels:	1
DB:	1	Gaps:	1
US-09-917-376-4 (1-89) x AF078038S5 (1-6005)			
Qy	4	GlyValLysValGlnTyrIlysAsnAspSerAlaProGlyAspAsnGlnIleLysPro	23
Db	2648	GGTGTGAAGGTACTGTACAAGAACCAATGACAAGTGGCAGCACAGGTTCTATAAGGCCG	2707
Qy	24	GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg	43
Db	2708	TGGTTTAAGATAGTGAATGGAGCGCAGCAGCAGTGTGTGATCTTAGCAGGGTTAAGATAGA	2767
Qy	44	TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaLa	63
Db	2768	TACTGTGTACACAGTGGTGTGACAAGCCACACAGAGTGGGTA--TGTGACTGGGCACAG	2824
Qy	64	MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla	83
Db	2825	ATVAGGGCAAGCAATGTGCATTCATTTGTGAAGCTTAGCAGCGGAGTGAGTGAGCG	2884
Qy	84	AspThrTyrLeuGln	88
Db	2885	GATTATTACCTGGAG	2899
RESULT 14			
E35100	E35100	6416 bp	DNA
LOCUS	E35100	Truncated cellulase composition.	linear
DEFINITION	E35100		PAT 18-JUN-2001
ACCESSION	E35100		
VERSION	E35100.1	GI:13018925	
KEYWORDS	JP 1999221086-A/2.		
SOURCE	unidentified		
ORGANISM	unclassified.		
REFERENCE	1 (bases 1 to 6416)		
AUTHORS	Paiji, A., Petaer, L.B., Roy, M.D., Gurahamu, K.F., Moreland, D.G., Hyu, M. and Dalian, P.W.		
TITLE	Truncated cellulase composition		
JOURNAL	Patent: JP 1999221086-A 2 17-AUG-1999;		
COMMENT	CLARIANT INTERNATIONAL LTD		
OS	Unknown		
PN	JP 1999221086-A/2		
PD	17-AUG-1999		
PF	21-SEP-1998	JP 1998283606	
PR	19-SEP-1997	US 08/932571	
PI	PALJI ANDERSON, PETAER L BAGUKUISUTO, ROY M DANIEL, PI		
GURAHAMU K FARINOTD,			
PI MORELAND DAVID GIBUSU, HYU MORGAN, DAIAN PURATONOTISU WILLIAM			
PC C112N15/09, C11D3/386, C12N1/21, C12N9/42// (C12N1/21, C12R1:19) , PC			
PC (C12N9/42, C12R1:19) , C12N15/00			
CC			
PH	Key	Location/Qualifiers	
FT	source	1..6416	
FT		/organism='Unknown'.	
FEATURES	Location/Qualifiers		
source	1..6416		
	/organism="unidentified"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:32644"		
ORIGIN			



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 12, 2004, 03:07:10 ; Search time 123.455 Seconds  
(without alignments)  
3062.577 Million cell updates/sec

Title: US-09-917-376-4

Perfect score: 470

Sequence: 1 VSGGVQVYKNDSPGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+\_p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US09917376/runat\_11052004\_114534\_28231/app\_query.fasta\_1.1429  
-DB=N Geneseq\_29Jan04 -QMT=fastcap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blowsum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MOD=LOCAL -OUTFWT=pro -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09917376@cgn\_1\_1758@runat\_11052004\_114534\_28231 -NCFU=6 -ICPU=3  
-NO WMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq\_29Jan04:\*

- 1: geneseqn1980s:\*
- 2: geneseqn1990s:\*
- 3: geneseqn2000s:\*
- 4: geneseqn2001as:\*
- 5: geneseqn2001bs:\*
- 6: geneseqn2002as:\*
- 7: geneseqn2003as:\*
- 8: geneseqn2003bs:\*
- 9: geneseqn2003cs:\*
- 10: geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	469	99.8	2289	7	ABZ77633 Nucleotid
2	466	99.1	3687	7	ABZ77634 Nucleotid
3	462	98.3	2869	7	ABZ77632 Nucleotid
4	462	98.3	2869	9	ADD22922 Acidothe
5	436	92.8	3365	7	ABZ76162 A. cellu
6	231.5	49.3	2600	2	AAQ15178 Portion o
7	195.5	41.6	2029	2	AAZ55660 DNA seque
8	195.5	41.6	2029	6	AAD26568 Active ce

9	194.5	41.4	6415	2	AAZ55662	Aax55662 DNA seque
10	194.5	41.4	6416	6	AAD26526	Aad26526 Active ce
11	192.5	41.0	11706	2	AAZ55661	Aax55661 DNA seque
12	192.5	41.0	11707	6	AAD26525	Aad26525 Active ce
13	188	40.0	1434	6	AAL41028	Aal41028 CMCase ge
14	188	40.0	1488	6	AAL41025	Aal41025 CMCase ge
15	188	40.0	2510	6	ABK53202	Abk53202 Bacillus
16	186	39.6	2175	2	AAQ49820	Aaq49820 NK-1 cell
17	182.5	38.8	2977	2	AAQ13001	Aaq13001 Endo1 gen
18	165	35.1	1438	3	AZ45336	Az45336 DNA encod
19	165	35.1	1482	2	AAZ30978	Aaz30978 DNA encod
20	165	35.1	1482	2	AAZ31562	Aaz31562 Pectate l
21	165	35.1	5562	2	AAT86625	Ata86625 C. thermo
22	157	33.4	1314	6	ABK73393	Abk73393 Bacillus
23	155.5	33.1	4059	5	AAH75059	Aah75059 Nucleotid
24	150.5	32.0	4056	5	AAH75060	Aah75060 Nucleotid
25	130.5	27.8	4452	7	ACA27429	Aca27429 Prokaryot
26	126.5	26.9	1303	9	ADC27473	Adc27473 Fusion pr
27	126.5	26.9	1747	9	ADC27474	Adc27474 Fusion pr
28	126.5	26.9	3489	9	ADC27475	Adc27475 Fusion pr
29	120.5	25.6	1775	2	AAQ13003	Aaq13003 Endo3 gen
30	114	24.3	486	2	AAQ72917	Aaq72917 Cellulose
31	114	24.3	486	2	AAV74072	Aav74072 C. cellul
32	114	24.3	499	2	AAZ4930	Aax24930 Clostridi
33	114	24.3	507	5	AAD11042	Aad11042 Clostridi
34	114	24.3	768	6	AAF86248	Aaf86248 DNA seque
35	114	24.3	768	6	ABK52403	Abk52403 E. coli c
36	114	24.3	984	5	AAD11046	Aad11046 Chimeric
37	114	24.3	1030	5	AAD11044	Aad11044 Clostridi
38	114	24.3	1288	5	AAD11045	Aad11045 Clostridi
39	113	24.0	573	5	AAD11043	Aad11043 Clostridi
40	113	24.0	1146	2	AAZ27550	Aaz27550 Gaussia l
41	113	24.0	1146	6	AAD22201	Aad22201 Gaussia s
42	78.5	16.7	764	3	AAF12881	Aaf12881 Aspergill
43	78.5	16.7	3147	7	ACA25846	Aca25846 Prokaryot
44	77	16.4	8107	3	AAZ92616	Aaz92616 Genomic D
45	76	16.2	96597	8	ADA02720	Ada02720 Human NFA

ALIGNMENTS

RESULT 1	ABZ77633	standard; DNA; 2289 BP.
ID	ABZ77633	standard; DNA; 2289 BP.
XX	AC	ABZ77633;
XX	DT	03-JUN-2003 (first entry)
XX	DE	Nucleotide sequence of the ManA polypeptide.
XX	KW	ManA; mannanase A; glycoside hydrolase; enzyme; hemicellulose; sugar;
XX	KW	food; feed; paper pulp; biofuel; mannanase; gene; ss.
XX	OS	Acidothermus cellulolyticus.
XX	FT	Key Location/Qualifiers
XX	FT	CDS 1..2289
XX	FT	/*tag= a
XX	FT	/product= "ManA"
XX	PN	WO2003012110-A1.
XX	PD	13-FEB-2003.
XX	PF	28-JUL-2001; 2001WO-US023819.
XX	PR	28-JUL-2001; 2001WO-US023819.
XX	PA	(MIDE ) MIDWEST RES INST.
XX	PI	Ding S, Adney WS, Vinzant TB, Himmel ME;

```
DR WPI; 2003-248182/24.
DR P-PSDB; ABP73022.
XX
XX Novel thermal tolerant mannanase A polypeptide derived from Acidothermus
PT cellulolyticus, useful for reducing hemicellulose in a starting material,
PT for processing of food, and as bulking agents in food stuffs.
XX
XX Example 1; Page 23; 46pp; English.
XX
XX The present sequence encodes ManaA, a thermostable mannanase A polypeptide
CC derived from Acidothermus cellulolyticus. ManaA is a member of the
CC glycoside hydrolase family of enzymes. ManaA is useful for reducing
CC hemicellulose in a starting material to simpler carbohydrate units, and
CC ultimately to sugars which are useful in the food, feed, paper pulp, and
CC biofuels industries. It is useful for the processing of food and in food
CC stuffs as bulking agents, and for the degradation of mannanase. ManaA is
CC also useful to raise polyclonal and monoclonal antibodies that are useful
CC in purifying ManaA, or detecting ManaA polypeptide expression, and as well
CC as reagent tools for characterizing the molecular actions of ManaA
CC polypeptides
XX
XX Sequence 2289 BP; 463 A; 700 C; 688 G; 438 T; 0 U; 0 Other;
SQ
Alignment Scores:
Pred. No.: 4.13e-48 Length: 2289
Score: 469.00 Matches: 88
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.79% Indels: 0
DB: 7 Gaps: 0
US-09-917-376-4 (1-89) x ABZ77633 (1-2289)
QY 1 ValSerGlyGlyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
DB 1363 GTGTCGGGTGGGTGAAGTGCAGTACAGAACAAATGATTCGGCCGGGTGATACACG 1422
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
DB 1423 ATCAAAACCGGGTCTCCAGTTGGTGAATACGGGGTCTGTCGTCGGTGAATTTGTCGACGGTG 1482
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
DB 1483 ACGGTGCGGTACTGTTTACCCGGGATGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1542
QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
DB 1543 TGGGGCGCGATGGGGTGGGGAATATCCGCGCTTGTTCGGTTCGGTGAACCCGGCGACG 1602
QY 81 ProThrAlaAspThrTyrLeuGln 88
DB 1603 CCGACGGCGGACACCTACCTGCAG 1626
RESULT 2
ID ABZ77634
XX ABZ77634 standard; DNA; 3687 BP.
XX
XX AC ABZ77634;
XX
XX 03-JUN-2003 (first entry)
XX
XX Nucleotide sequence of the GuxA polypeptide.
XX
XX GuxA; cellulase; glycoside hydrolase; enzyme; cellulose; sugar; biofuel;
KW detergent; pulp processing; paper processing; feed processing; textile;
KW gene; ss.
XX
XX Acidothermus cellulolyticus.
XX
XX Key Location/Qualifiers
FT CDS 1..3687
FT /*tag= a
FT /product= "GuxA"
```





DB: 9 Gaps: 0

US-09-917-376-4 (1-89) x ADD22922 (1-2869)

QY 1 ValSerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20  
 Db 2605 GTGTCGGTGGGTGAGGTGAGTATAGATATGATTCGGCGCGGTGATATCAG 2664

QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40  
 Db 2665 ATCAAGCCGGGTTTCAGCGTGGTGAATACCGGGTCGTCGGTGGATTTGTCGACGGTG 2724

QY 41 ThrValArgTyrTyrPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
 Db 2725 ACGGTGCGGTACTGTTTACCGCGGATGGTGGTCGTCGACACTGGTGACACTGTGAC 2784

QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80  
 Db 2785 TGGCGCGGATCGGTGTGGGATATCCGGCTCGTTCGGTCCGGTGAACCGCGGACG 2844

QY 81 ProThrAlaAspThrTyrLeuGln 88  
 Db 2845 CCGACGGCGGACACCTACCTGCAG 2868

RESULT 5  
 ABZ76162  
 ID ABZ76162 standard; DNA; 3365 BP.  
 XX  
 AC ABZ76162;  
 DT 29-MAY-2003 (first entry)  
 XX  
 DE A. cellulolyticus Gux1 protein encoding DNA.  
 KW Gux1; thermal tolerant; exoglucanase; glycoside hydrolase; cellulose;  
 KW biofuel; detergent; pulp; paper processing; feed processing; textile;  
 KW cellulase; gene; ds.  
 OS  
 XX Acidothermus cellulolyticus.  
 FH Key Location/Qualifiers  
 FT CDS 1..3365  
 FT /tag= a  
 FT /product= "Gux1"  
 FT /transl\_except= (pos: 682..683, aa: Pro)  
 FT /note= "this codon has an apparent one nucleotide  
 FT basepair deletion which alters the reading frame"  
 XX  
 PN WO2003012095-A1.  
 XX  
 XX  
 PD 13-FEB-2003.  
 XX  
 PF 28-JUL-2001; 2001WO-US023820.  
 PR 28-JUL-2001; 2001WO-US023820.  
 XX  
 XX (MIDE ) MIDWEST RES INST.  
 XX  
 PI Adney WS, Ding S, Vinzant TB, Himmel ME, Decker SR;  
 PI Lantz Mccarter S;  
 XX  
 XX WPI; 2003-300494/29.  
 DR P-PSDB; ABP71656.  
 XX  
 PT New thermal tolerant Gux1 peptide having specified amino acid sequence,  
 PT useful in the degradation of cellulose to biofuels.  
 XX  
 PS Disclosure; Page 22-23; 44pp; English.  
 XX  
 CC The invention relates to a thermal tolerant Gux1 peptide from A.  
 CC cellulolyticus. The Gux1 exoglucanase is a member of the glycoside  
 CC hydrolase family and comprises a catalytic domain GH48, carbohydrate  
 CC binding domain type III, and a carbohydrate binding domain type II. The

CC polypeptide is useful in the degradation of cellulose into biofuel, or  
 CC for conversion of biomass to biofuel additives. It is used in detergents,  
 CC pulp and paper processing, food and feed processing, and in textile  
 CC processing. It can also be used alone or in combination with other  
 CC cellulase or glycoside hydrolases. The novel polypeptide generates  
 CC alternative cellulase enzymes capable of assisting in the commercial-  
 CC scale processing of cellulose to sugar for use in biofuel production. The  
 CC present sequence represents a A. cellulolyticus Gux1 cellulase encoding  
 CC DNA  
 XX  
 SQ Sequence 3365 BP; 669 A; 1117 C; 1041 G; 538 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 8.56e-44 Length: 3365  
 Score: 436.00 Matches: 81  
 Percent Similarity: 97.62% Conservative: 1  
 Best Local Similarity: 96.43% Mismatches: 2  
 Query Match: 92.77% Indels: 0  
 DB: 7 Gaps: 0

US-09-917-376-4 (1-89) x ABZ76162 (1-3365)

QY 5 ValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysProGly 24  
 Db 112 CTCAAAGCGGAGTATAGAACAAATGATTCGGCGCGGAGTGACCAACGATCAACCGGGT 171

QY 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArgTyr 44  
 Db 172 CTCAGTGTGGTGAATACCGGGTCGTCGGTGGATTTGTCGACGGTGACGGTGGCGTAC 231

QY 45 TrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAlaMet 64  
 Db 232 TGGTTTCCCGCGGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 291

QY 65 GlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAlaAsp 84  
 Db 292 GGGTGTGGGAAATATCCGGCTCGTTCGGTGGTGAACCGCGGACCCCGCGGAC 351

QY 85 ThrTyrLeuGln 88  
 Db 352 ACCTACTGCAG 363

RESULT 6  
 AAQ15178  
 ID AAQ15178 standard; DNA; 2600 BP.  
 XX  
 AC AAQ15178;  
 XX  
 DT 27-AUG-2003 (revised)  
 DT 25-MAR-2003 (revised)  
 DT 17-MAR-1992 (first entry)  
 XX  
 DE Portion of pAEC-1 contg. cellulase AE-1 gene.  
 XX  
 KW Detergents; pharmaceuticals; deinking; carboxymethylcellulose; ss.  
 XX Aeromonas sp.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 192..2540  
 FT /tag= a  
 FT /product= "cellulase AE-1"  
 XX  
 PN JP03251174-A.  
 XX  
 PD 08-NOV-1991.  
 XX  
 PF 28-FEB-1990; 90JP-00045465.  
 XX  
 PR 28-FEB-1990; 90JP-00045465.  
 XX  
 PA (OJIP ) OJI PAPER CO.  
 XX

DR WPI; 1991-373412/51.  
 DR P-PSDB; AAR15625.  
 XX  
 PT Cellulase AE-1 for e.g. mfr. of pharmaceuticals and foodstuffs - of opt.  
 PT pH when carboxymethylcellulose is used as substrate.  
 XX  
 PS Claim 3; Fig 2; 8pp; Japanese.  
 XX  
 XX The plasmid, pAEC 1, was prepd. ligating chromosomal DNA contg. the gene  
 CC (obtd. from Aeromonas) into pUC18. The gene can be used to produce  
 CC recombinant enzyme which is used for the effective utilis- ation of  
 CC biomass resources and the mfr. of pharmaceuticals and foodstuffs, and  
 CC also for the detergent and deinking of waste paper. (Updated on 25-MAR-  
 XX 2003 to correct PA field.) (Updated on 27-AUG-2003 to correct OS field.)  
 XX  
 SQ Sequence 2600 BP; 687 A; 649 C; 700 G; 564 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 1.62e-18 Length: 2600  
 Score: 231.50 Matches: 42  
 Percent Similarity: 66.67% Conservative: 16  
 Best Local Similarity: 48.28% Mismatches: 28  
 Query Match: 49.28% Indels: 1  
 DB: 2 Gaps: 1

US-09-917-376-4 (1-89) x AAK15178 (1-2600)

QY 2 SerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 21  
 Db 2082 AGCGGGGATTTGGCGTTTCAGTACAAAACCGGAGATCGAATGCGCCGACACAGTTC 2141

QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41  
 Db 2142 AAGCTCTATTCAATATCGTGAACAGGGGCTGGCAGTGCCTTTAAGCAGCTGAGT 2201

QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61  
 Db 2202 CTGCGATATTATTTTCACAGCTGAC---GGCAATGACCAACTGCAATACAAATTTGCTACTGG 2258

QY 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81  
 Db 2259 GCCATGTGTAGGTCTCAACCTGAACGGGGCTTTTCGTGAAAATGAATCCGGGCAAGCG 2318

QY 82 ThrAlaAspThrTyrLeuGln 88  
 Db 2319 AACGCCGATACCTACTCTGGAG 2339

RESULT 7  
 ID AAX55660  
 AC AAX55660;  
 XX  
 XX 30-JUL-1999 (first entry)  
 DE DNA sequence encoding truncated cellulase Cel E3/B5.  
 XX Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;  
 KW Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;  
 KW cotton-containing fabric; stonewashing; ss.  
 XX Unidentified.  
 OS  
 XX EP921188-A2.  
 PN  
 XX 09-JUN-1999.  
 PD  
 XX 15-SEP-1998; 98EP-00810919.  
 PF  
 XX 19-SEP-1997; 97US-00932571.  
 PR  
 XX (CLRN ) CLARIANT FINANCE BVI LTD.  
 PA  
 XX

PI Anderson P., Bergquist PL, Daniels RM, Farrington GK, Gibbs MD;  
 PI Morgan H., Williams DP;  
 XX  
 DR WPI; 1999-315403/27.  
 DR P-PSDB; AAV13494.  
 XX  
 PT New truncated cellulase proteins, useful in detergents and for producing  
 PT 'stonewashed' denim.  
 XX  
 PS Disclosure; Page 41-42; 65pp; English.  
 XX  
 XX The invention relates to a recombinant cellulase active protein free of  
 CC proteinases of native thermophilic and alkalophilic origin, comprising  
 CC the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3,  
 CC Cel 6 or Cel E3/B5, or a stability region from one of the defined full-  
 CC length sequences, or functional equivalents. Cel B5 extends from amino  
 CC acid A1011 to P1424 or K1425 or N1426, and Cel B4/5 extends from amino  
 CC acid K635 to N1426 in the sequence shown in AAY13492; Cel E1 extends from  
 CC amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel E1/2/3  
 CC extends from Y39 to G812, Cel E6 extends from amino acid V1233 to K1751  
 CC and the stability region extends from amino acid E482 to G635 in the  
 CC sequence shown in AAY13493; Cel E3/B5 is shown in AAY13494. The new  
 CC enzymes are useful in laundry detergent compositions to prevent or remove  
 CC staining, backstaining or graying, for use on cellulosic materials  
 CC including cotton-containing fabrics. They are especially useful for  
 CC preventing redeposition of colorant during stonewashing, and for  
 CC processing of textiles where cellulose breakdown is required. The new  
 CC truncated enzymes show reduced redeposition of dye compared to using non-  
 CC truncated cellulase compositions

Sequence 2029 BP; 650 A; 329 C; 520 G; 530 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 3.53e-14 Length: 2029  
 Score: 195.50 Matches: 38  
 Percent Similarity: 60.23% Conservative: 15  
 Best Local Similarity: 43.18% Mismatches: 34  
 Query Match: 41.60% Indels: 1  
 DB: 2 Gaps: 1

US-09-917-376-4 (1-89) x AAX55660 (1-2029)

QY 1 ValSerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20  
 Db 1 ATGGGAAGTGGTGTGAAGGTACTGTACAGACAAATGACACAGTGCACAGGTCTCT 60

QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40  
 Db 61 ATAAGGCCGCTGGTTTAAGATAGTAGTAATGGAGGCGACGAGTGTGTGATCTTAGCAGGTT 120

QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
 Db 121 AAGATAAGATACCTGTGTACACAGTGGATGGTGTGTGACACAGCCACAGAGTGGCGTA---TGTGAC 177

QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80  
 Db 178 TGGGCACAGATAGGGCCCAAGCAATGTACATTCATTTTGTGAGCTTAGCAGCGAGTG 237

QY 81 ProThrAlaAspThrTyrLeuGln 88  
 Db 238 AGTGGAGCGGATATTACCTGGAG 261

RESULT 8  
 AAD26568  
 ID AAD26568 standard; DNA; 2029 BP.  
 XX  
 XX AAD26568;  
 AC  
 XX 26-MAR-2002 (first entry)  
 DT  
 XX Active cellulase hybrid protein, E3/B5 DNA.  
 DE  
 XX Active cellulase protein; alkalophilic; textile processing; proteinase;  
 KW

KW detergent additive; stonewashed appearance; cotton-containing denim;  
 KW CelB5; thermophilic; commercial detergent; E3/B5 hybrid protein; ds.  
 XX Unidentified.  
 OS  
 XX  
 XX Key Location/Qualifiers  
 FT 1..1851  
 FT CDS  
 FT /\*tag= a  
 FT /product= "E3/B5 hybrid protein"  
 XX  
 XX US6294366-B1.  
 XX  
 XX 25-SEP-2001.  
 XX  
 XX 19-AUG-1998; 98US-00136574.  
 XX  
 XX 19-SEP-1997; 97US-00932571.  
 XX  
 XX (CLRN ) CLARIANT FINANCE BVI LTD.  
 XX  
 XX Farrington GK, Anderson P, Bergquist P, Daniels R, Gibbs MD;  
 PI Morgan H, Williams DP;  
 XX  
 XX WPI: 2002-081780/11.  
 DR P-PSDB; AAE16325.  
 XX  
 XX New cellulase active protein, useful in textile processing or commercial  
 PT detergents, e.g. for improving the feel or appearance of cotton-  
 PT containing fabrics, is stable under conditions of alkaline pH and  
 PT elevated temperatures.  
 XX  
 XX Disclosure; Col 71-74; 61pp; English.  
 XX  
 XX The present invention relates to a cellulase active protein, which is  
 CC substantially free of proteinases of native thermophilic and  
 CC alkaliphilic origin, where the cellulase active protein consists of the  
 CC CelB5 amino acid sequence. The cellulase active protein is useful for  
 CC treating cellulosic materials including cotton-containing fabrics, as  
 CC detergent additives. The cellulase active protein is also useful for  
 CC improving the feel and/or appearance of cotton-containing fabrics, for  
 CC removing surface fibers from cotton-containing knits or for imparting  
 CC stonewashed appearance to cotton-containing denims. The present proteins  
 CC are stable under condition of alkaline pH and elevated temperatures, thus  
 CC suitable for textile processing and in commercial detergents. The present  
 CC sequence is E3/B5 hybrid protein DNA  
 XX  
 XX Sequence 2029 BP; 650 A; 329 C; 520 G; 530 T; 0 U; 0 Other;  
 SQ  
 Alignment Scores:  
 Pred. No.: 3.53e-14 Length: 2029  
 Score: 195.50 Matches: 38  
 Percent Similarity: 60.23% Conservative: 15  
 Best Local Similarity: 43.18% Mismatches: 34  
 Query Match: 41.60% Indels: 1  
 DB: 6 Gaps: 1  
 US-09-917-376-4 (1-89) x AAD26568 (1-2029)  
 QY 1 ValSerGlyValValValGlnThrValSerValAspSerValProGlyAspAsnGln 20  
 DB 1 ATGGGAAGTGGTGTGAAGTACTGTACAGAACATGACAGCAAGTGCAGACAGGTTCT 60  
 QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40  
 DB 61 ATAGGCGGTGGTTTAAGTAGTAGTATGAGGAGCAGCAGGATGTTGATTTAGCAGGTT 120  
 QY 41 ThrValArgTrpPheThrArgAspGlyGlySerSerThrLeuValThrAsnCysAsp 60  
 DB 121 AAGATAAGATACCTGGTACACAGCTGGATGGTGTGCAAGCCACAGAGTCCGCTA---TGTGAC 177  
 QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80  
 DB 178 TGGGCACAGATAGGGGCAAGCAATGTGACATTTCAATTTTGTGAAGCTTAGCAGCGAGTG 237

QY 81 ProThrAlaAspThrTyrLeuGln 88  
 DB 238 AGTGGAGCGGATTATTACCTGGAG 261  
 RESULT 9  
 AAX55662  
 ID AAX55662 standard; DNA; 6415 BP.  
 XX  
 XX AAX55662;  
 XX  
 XX 30-JUL-1999 (first entry)  
 XX  
 XX DNA sequence encoding truncated cellulases.  
 XX  
 XX Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;  
 KW Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;  
 KW cotton-containing fabric; stonewashing; ss.  
 XX  
 XX Unidentified.  
 OS  
 XX EP921188-A2.  
 XX  
 XX 09-JUN-1999.  
 XX  
 XX 15-SEP-1998; 98EP-00810919.  
 XX  
 XX 19-SEP-1997; 97US-00932571.  
 XX  
 XX (CLRN ) CLARIANT FINANCE BVI LTD.  
 XX  
 XX Anderson P, Bergquist PL, Daniels RM, Farrington GK, Gibbs MD;  
 PI Morgan H, Williams DP;  
 XX  
 XX WPI: 1999-315403/27.  
 DR P-PSDB; AAY13493.  
 XX  
 XX New truncated cellulase proteins, useful in detergents and for producing  
 PT 'stonewashed' denim.  
 XX  
 XX Disclosure; Page 24-25; 65pp; English.  
 XX  
 XX The invention relates to a recombinant cellulase active protein free of  
 CC proteinases of native thermophilic and alkaliphilic origin, comprising  
 CC the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3,  
 CC Cel 6 or Cel E3/B5, or a stability region from one of the defined full-  
 CC length sequences, or functional equivalents. Cel B5 extends from amino  
 CC acid A1011 to F1424 or K1425 or N1426, and Cel B4/5 extends from amino  
 CC acid K635 to N1426 in the sequence shown in AAY13492; Cel E1 extends from  
 CC amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel E1/2/3  
 CC extends from Y39 to G812, Cel E6 extends from amino acid V1233 to K1751  
 CC and the stability region extends from amino acid E482 to G635 in the  
 CC sequence shown in AAY13493; Cel E3/B5 is shown in AAY13494. The new  
 CC enzymes are useful in laundry detergent compositions to prevent or remove  
 CC staining, backstaining or graying, for use on cellulosic materials  
 CC including cotton-containing fabrics. They are especially useful for  
 CC preventing redeposition of colorant during stonewashing, and for  
 CC processing of textiles where cellulose breakdown is required. The new  
 CC truncated enzymes show reduced redeposition of dye compared to using non-  
 CC truncated cellulase compositions  
 XX  
 XX Sequence 6415 BP; 2067 A; 1082 C; 1689 G; 1541 T; 0 U; 36 Other;  
 SQ  
 Alignment Scores:  
 Pred. No.: 2.05e-13 Length: 6415  
 Score: 194.50 Matches: 38  
 Percent Similarity: 61.18% Conservative: 14  
 Best Local Similarity: 44.71% Mismatches: 32  
 Query Match: 41.38% Indels: 1  
 DB: 2 Gaps: 1  
 US-09-917-376-4 (1-89) x AAX55662 (1-6415)

QY 4 GlyVallyValGlnTyrIlyAsnAspSerAlaProGlyAspAsnGlnIleIysPro 23  
 DB 2664 GGTGTGAAGTACTGTACAAGAACATGACACAAAGTGGCAGCACAGGTTCTATAAGGCG 2723  
 QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43  
 DB 2724 TGGTTTAAGATAGTAAGTGAAGGAGGAGCAGCAAGTGTGATCTTACGAGGTTAAGATAAGA 2783  
 QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63  
 DB 2784 TACTGGTACACAGTGGATGGTGCACAGCCACAGAGTGGCGTA--TGTGACTGGGCACAG 2840  
 QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83  
 DB 2841 ATAGGGGCAAGCAATGTGACATTCATTTTGTGAAGCTTAGCAGCGGAGTGAGTGGAGCG 2900  
 QY 84 AspThrTyrLeuGln 88  
 DB 2901 GATTATTACCTGGAG 2915  
 RESULT 10  
 AAD26526  
 ID AAD26526 standard; DNA; 6416 BP.  
 AC AAD26526;  
 DT 26-MAR-2002 (first entry)  
 DE Active cellulase protein, cele gene.  
 XX  
 KW Active cellulase protein; alkalophilic; textile processing; proteinase;  
 KW detergent additive; stonewashed appearance; cotton-containing denim;  
 KW CelB5; thermophilic; commercial detergent; cele gene; ds.  
 XX  
 OS Unidentified.  
 FH Key Location/Qualifiers  
 FT CDS 634..5889  
 FT /\*tag= a  
 FT /product= "cele protein"  
 FT misc\_feature 748..2538  
 FT /\*tag= c  
 FT /product= "DNA encoding E1/2 protein"  
 FT misc\_feature 748..2076  
 FT /\*tag= b  
 FT /product= "DNA encoding E1 protein"  
 XX  
 PN US6294366-B1.  
 XX  
 PD 25-SEP-2001.  
 XX  
 PF 19-AUG-1998; 98US-00136574.  
 XX  
 PR 19-SEP-1997; 97US-00932571.  
 XX  
 PA (CLRN ) CLARIANT FINANCE BVI LTD.  
 XX  
 PI Farrington GK, Anderson P, Bergquist P, Daniels R, Gibbs MD;  
 PI Morgan H, Williams DP;  
 XX  
 DR WPI; 2002-081780/11.  
 DR P-PSDB; AAE16324.  
 XX  
 PT New cellulase active protein, useful in textile processing or commercial  
 PT detergents, e.g. for improving the feel or appearance of cotton-  
 PT containing fabrics, is stable under conditions of alkaline pH and  
 PT elevated temperatures.  
 XX  
 PS Disclosure; Col 37-44; 6lpp; English.  
 XX  
 CC The present invention relates to a cellulase active protein, which is  
 CC substantially free of proteinases of native thermophilic and  
 CC alkalophilic origin, where the cellulase active protein consists of the

CC CelB5 amino acid sequence. The cellulase active protein is useful for  
 CC treating cellulosic materials including cotton-containing fabrics, as  
 CC detergent additives. The cellulase active protein is also useful for  
 CC improving the feel and/or appearance of cotton-containing fabrics, for  
 CC removing surface fibers from cotton-containing knits or for imparting  
 CC stonewashed appearance to cotton-containing denims. The present proteins  
 CC are stable under condition of alkaline pH and elevated temperatures, thus  
 CC suitable for textile processing and in commercial detergents. The present  
 CC sequence is cele gene  
 XX  
 SQ Sequence 6416 BP; 2068 A; 1082 C; 1689 G; 1541 T; 0 U; 36 Other;  
 Alignment Scores:  
 Pred. No.: 2.05e-13 Length: 6416  
 Score: 194.50 Matches: 38  
 Percent Similarity: 61.18% Conservative: 14  
 Best Local Similarity: 44.71% Mismatches: 32  
 Query Match: 41.38% Indels: 1  
 DB: 6 Gaps: 1  
 US-09-917-376-4 (1-89) x AAD26526 (1-6416)  
 QY 4 GlyVallyValGlnTyrIlyAsnAspSerAlaProGlyAspAsnGlnIleIysPro 23  
 DB 2665 GGTGTGAAGTACTGTACAAGAACATGACACAAAGTGGCAGCACAGGTTCTATAAGGCG 2724  
 QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43  
 DB 2725 TGGTTTAAGATAGTAAGTGAAGGAGGAGCAGCAAGTGTGATCTTACGAGGTTAAGATAAGA 2784  
 QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63  
 DB 2785 TACTGGTACACAGTGGATGGTGCACAGCCACAGAGTGGCGTA--TGTGACTGGGCACAG 2841  
 QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83  
 DB 2842 ATAGGGGCAAGCAATGTGACATTCATTTTGTGAAGCTTAGCAGCGGAGTGAGTGGAGCG 2901  
 QY 84 AspThrTyrLeuGln 88  
 DB 2902 GATTATTACCTGGAG 2916  
 RESULT 11  
 AAX55661  
 ID AAX55661 standard; DNA; 11706 BP.  
 AC AAX55661;  
 DT 30-JUL-1999 (first entry)  
 DE DNA sequence encoding truncated cellulases Cel B4/5 and Cel B5.  
 XX  
 KW Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;  
 KW Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;  
 KW cotton-containing fabric; stonewashing; ss.  
 XX  
 OS Unidentified.  
 XX  
 PN EP921188-A2.  
 XX  
 PD 09-JUN-1999.  
 XX  
 PF 15-SEP-1998; 98EP-00810919.  
 XX  
 PR 19-SEP-1997; 97US-00932571.  
 XX  
 PA (CLRN ) CLARIANT FINANCE BVI LTD.  
 XX  
 PI Anderson P, Bergquist PL, Daniels RM, Farrington GK, Gibbs MD;  
 PI Morgan H, Williams DP;  
 XX  
 DR WPI; 1999-315403/27.  
 DR P-PSDB; AAV13492.

XX New truncated cellulase proteins, useful in detergents and for producing  
PT 'stonewashed' denim.  
PS Disclosure; Page 20-23; 65pp; English.  
XX The invention relates to a recombinant cellulase active protein free of  
CC proteinases of native thermophilic and alkalophilic origin, comprising  
CC the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3,  
CC Cel 6 or Cel E3/B5, or a stability region from one of the defined full-  
CC length sequences, or functional equivalents. Cel B5 extends from amino  
CC acid A1011 to P1424 or K1425 or N1426, and Cel B4/5 extends from amino  
CC acid K635 to N1426 in the sequence shown in AAY13492; Cel E1 extends from  
CC amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel E1/2/3  
CC extends from Y39 to G812, Cel B6 extends from amino acid V1233 to K1751  
CC and the stability region extends from amino acid E482 to G635 in the  
CC sequence shown in AAY13493; Cel E3/B5 is shown in AAY13494. The new  
CC enzymes are useful in laundry detergent compositions to prevent or remove  
CC staining, backstaining or graying, for use on cellulosic materials  
CC including cotton-containing fabrics. They are especially useful for  
CC preventing redeposition of colorant during stonewashing, and for  
CC processing of textiles where cellulose breakdown is required. The new  
CC truncated enzymes show reduced redeposition of dye compared to using non-  
CC truncated cellulase compositions  
XX  
SQ Sequence 11706 BP; 3828 A; 1994 C; 2994 G; 2890 T; 0 U; 0 Other;  
  
Alignment Scores:  
Pred. No.: 7, 82e-13 Length: 11706  
Score: 192.50 Matches: 37  
Percent Similarity: 62.35% Conservatives: 16  
Best Local Similarity: 43.53% Mismatches: 31  
Query Match: 40.96% Indels: 1  
DB: 2 Gaps: 1  
  
US-09-917-376-4 (1-89) x AAX55661 (1-11706)  
QY 4 GlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23  
DB 4038 GGTTCGAGGTACTATACAAAGCAATGAGCAAGTCGAGCAAGTCTTATAAGCGC 4097  
QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43  
DB 4098 TGGTTTAAAGATAGTGAATGAGGAGCAGCAGCGTGTGATCTTAGCAGGGTTAAGATAAGA 4157  
QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAla 63  
DB 4158 TACTGTACACAGTGTGATGTCAGCAAGCCACAGAGTGGCGTA---TGTGACTGGGCACAG 4214  
QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83  
DB 4215 ATAGGGCAAGCAATGTGACATTCATTTTGTGAGCTGAGCGGAGTGGAGCG 4274  
QY 84 AspThrTyrLeuGln 88  
DB 4275 GATTATTACTTGGAG 4289  
  
RESULT 12  
AAD26525  
ID AAD26525 standard; DNA; 11707 BP.  
XX  
AC AAD26525;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Active cellulase protein, celB gene.  
XX  
KW Active cellulase protein; alkalophilic; textile processing; proteinase;  
KW detergent additive; stonewashed appearance; cotton-containing denim;  
KW CelB5; thermophilic; commercial detergent; celB gene; ds.  
XX  
OS Unidentified.  
XX

FH Key Location/Qualifiers  
CDS 6255..10535  
FT /\*tag= a  
FT /product= "CelB protein"  
FT 8501..10532  
FT /\*tag= b  
FT /product= "DNA encoding B4/5 protein"  
FT 9255..10526  
FT /\*tag= c  
FT /product= "DNA encoding B/5 protein"  
XX  
PN US6294366-B1.  
XX  
PD 25-SEP-2001.  
XX  
PF 19-AUG-1998; 98US-00136574.  
XX  
PR 19-SEP-1997; 97US-00932571.  
XX  
PA (CLRN ) CLARIANT FINANCE BVI LTD.  
XX  
PI Farrington GK, Anderson P, Bergquist P, Daniels R, Gibbs MD;  
PI Morgan H, Williams DP;  
XX  
DR WPI; 2002-081780/11.  
DR P-PSDB; AAE16323.  
XX  
PT New cellulase active protein, useful in textile processing or commercial  
PT detergents, e.g. for improving the feel or appearance of cotton-  
PT containing fabrics, is stable under conditions of alkaline pH and  
PT elevated temperatures.  
XX  
PS Disclosure; Col 27-38; 61pp; English.  
XX  
CC The present invention relates to a cellulase active protein, which is  
CC substantially free of proteinases of native thermophilic and  
CC alkalophilic origin, where the cellulase active protein consists of the  
CC CelB5 amino acid sequence. The cellulase active protein is useful for  
CC treating cellulosic materials including cotton-containing fabrics, as  
CC detergent additives. The cellulase active protein is also useful for  
CC improving the feel and/or appearance of cotton-containing fabrics, for  
CC removing surface fibers from cotton-containing knits or for imparting  
CC stonewashed appearance to cotton-containing denims. The present proteins  
CC are stable under condition of alkaline pH and elevated temperatures, thus  
CC suitable for textile processing and in commercial detergents. The present  
CC sequence is celB gene  
XX  
SQ Sequence 11707 BP; 3827 A; 1995 C; 2994 G; 2891 T; 0 U; 0 Other;  
  
Alignment Scores:  
Pred. No.: 7, 83e-13 Length: 11707  
Score: 192.50 Matches: 37  
Percent Similarity: 62.35% Conservatives: 16  
Best Local Similarity: 43.53% Mismatches: 31  
Query Match: 40.96% Indels: 1  
DB: 6 Gaps: 1  
  
US-09-917-376-4 (1-89) x AAD26525 (1-11707)  
QY 4 GlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23  
DB 4038 GGTTCGAGGTACTATACAAAGCAATGAGCAAGTCGAGCAAGTCTTATAAGCGC 4097  
QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43  
DB 4098 TGGTTTAAAGATAGTGAATGAGGAGCAGCAGCGTGTGATCTTAGCAGGGTTAAGATAAGA 4157  
QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAla 63  
DB 4158 TACTGTACACAGTGTGATGTCAGCAAGCCACAGAGTGGCGTA---TGTGACTGGGCACAG 4214  
QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83  
DB 4215 ATAGGGCAAGCAATGTGACATTCATTTTGTGAGCTGAGCGGAGTGGAGCG 4274

```
Db 4215 ATAGGGGCAAGCAATGTGACATTCATTTTGTGAAGCTGAGCAGCGGAGTGAGTGAGCGG 4274
QY 84 AspThrTyrLeuGln 88
Db 4275 GATTATTACTTGGAG 4289
RESULT 13
ID AAL41028
XX AAL41028 standard; DNA; 1434 BP.
AC AAL41028;
XX 11-OCT-2002 (first entry)
DE CMCase gene with additional His encoding sequences.
KW Protein surface-displayed; genetic carrier; spore; monoclonal variant;
KW high-throughput screening; antibody; vaccine; enzyme; gene; ds.
XX Unidentified.
FH Key Location/Qualifiers
FT CDS 1..1434
FT /*tag= a
FT /product= "CMCase protein"
XX WO200255561-A1.
XX 18-JUL-2002.
XX 15-JAN-2002; 2002WO-KR000059.
XX 15-JAN-2001; 2001KR-00002156.
XX (GENO-) GENOFOCUS CO LTD.
XX Pan JG, Choi SK, Jung HC;
XX WPI; 2002-590661/63.
XX P-PSDB; AAO22444.
XX Methods for surface display of proteins, e.g. displaying a protein on the
XX surface of spores, improving a protein or isolating a substance, by
XX transforming a host cell harboring a genetic carrier with spore or virus
XX with the vector library.
XX Example 5; Page 109-113; 118pp; English.
XX The invention relates to methods for preparing a protein surface-
XX displayed on a genetic carrier, improving a protein or isolating a
XX substance in a mixture comprising transforming a host cell harbouring a
XX genetic carrier consisting of a spore or virus with the vector library.
XX The method is useful for surface display of proteins, particularly for
XX displaying a protein on the surface of e.g. spores, for improving a
XX protein, or for isolating a substance. These methods are useful in
XX obtaining monoclonal variants from a large library, or in high-throughput
XX screening of antibodies for use in therapy (e.g. as vaccines), diagnosis
XX or analysis. This polynucleotide sequence represents the CMCase gene with
XX additional His encoding sequences relating to the invention
XX Sequence 1434 BP; 493 A; 293 C; 341 G; 307 T; 0 U; 0 Other;
SQ
Alignment Scores:
Pred. No.: 1.94e-13 Length: 1434
Score: 188.00 Matches: 34
Percent Similarity: 61.18% Conservative: 18
Best Local Similarity: 40.00% Mismatches: 31
Query Match: 40.00% Indels: 2
DB: 6 Gaps: 1
US-09-917-376-4 (1-89) x AAL41028 (1-1434)
QY 4 GlyValIysValGlnTyrLysAsnAenAspSerAlaProGlyAspAsnGlnIleLysPro 23
```

---

```
Db 991 GGTATTTCGTCAATACAGACGAGGGGATGGGAGTATGAACAGCAACCAATCCGTCCG 1050
QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
Db 1051 CAGCTTCAATAAATAAATACGCAATACCAACGGTTGATTTAAAGATGTCACTGCCGT 1110
QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63
Db 1111 TACTGTATATACGCGAATAAAGCCAAACGTT-----GACTGTGACTACGCGAG 1164
QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
Db 1165 CTGGATGCGCAATGTGACATACAAGTTTGTGACGTGTGCATTAACCAACCAAGAGTGCA 1224
QY 84 AspThrTyrLeuGln 88
Db 1225 GATACCTATCTGGAA 1239
RESULT 14
ID AAL41025
XX AAL41025 standard; DNA; 1488 BP.
AC AAL41025;
XX 11-OCT-2002 (first entry)
DE CMCase gene with mutated signal sequence to enhance hydrophobicity.
KW Protein surface-displayed; genetic carrier; spore; monoclonal variant;
KW high-throughput screening; antibody; vaccine; enzyme; gene; ds.
XX Unidentified.
FH Key Location/Qualifiers
FT CDS 1..1491
FT /*tag= a
FT /product= "CMCase protein"
XX WO200255561-A1.
XX 18-JUL-2002.
XX 15-JAN-2002; 2002WO-KR000059.
XX 15-JAN-2001; 2001KR-00002156.
XX (GENO-) GENOFOCUS CO LTD.
XX Pan JG, Choi SK, Jung HC;
XX WPI; 2002-590661/63.
XX P-PSDB; AAO22443.
XX Methods for surface display of proteins, e.g. displaying a protein on the
XX surface of spores, improving a protein or isolating a substance, by
XX transforming a host cell harboring a genetic carrier with spore or virus
XX with the vector library.
XX Example 4; Page 98-103; 118pp; English.
XX The invention relates to methods for preparing a protein surface-
XX displayed on a genetic carrier, improving a protein or isolating a
XX substance in a mixture comprising transforming a host cell harbouring a
XX genetic carrier consisting of a spore or virus with the vector library.
XX The method is useful for surface display of proteins, particularly for
XX displaying a protein on the surface of e.g. spores, for improving a
XX protein, or for isolating a substance. These methods are useful in
XX obtaining monoclonal variants from a large library, or in high-throughput
XX screening of antibodies for use in therapy (e.g. as vaccines), diagnosis
XX or analysis. This polynucleotide sequence represents the CMCase gene with
XX a mutated signal sequence to enhance its hydrophobicity relating to the
XX invention
```

```
XX SQ Sequence 1488 BP; 501 A; 297 C; 357 G; 333 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 2,04e-13 Length: 1488
Score: 188.00 Matches: 34
Percent Similarity: 61.18% Conservatives: 18
Best Local Similarity: 40.00% Mismatches: 31
Query Match: 40.00% Indels: 2
DB: 6 Gaps: 1

US-09-917-376-4 (1-89) x AAL41025 (1-1488)
QY 4 GlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
DQ 1048 GGTATTCTGTACATAACAGCAGCGGATGGAGTATGACAGCAACCAATCCGTCGG 1107
QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
DQ 1108 CAGCTTCAATAAATAAATACCGCAATACCGGTTGATTAAAGATGCTACTGCCGT 1167
QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaLa 63
DQ 1168 TACTGGTATACCGCAAAACAAAGCCCAAAACGTT-----GACTGTGACTACGGCGAG 1221
QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
DQ 1222 CTTGGATCGGCAATGTGACATACAAAGTTTGTGACGTTCATAAACCAAGCAAGTGCA 1281
QY 84 AspThrTyrLeuGln 88
DQ 1282 GATACCTATCTGGAA 1296

RESULT 15
ID ABK53202 standard; DNA; 2510 BP.
AC ABK53202;
DT 12-AUG-2002 (first entry)
DE Bacillus subtilis spore coat protein cotG-CMCCase fusion gene.
KW Spore coat protein; spore surface; delivery system; bioconversion;
KW recombinant live vaccine; resistance property; protein microarray;
KW whole cell absorber; secretion signal; target signal; cell membrane;
KW surface display motif; cotG; CMCCase; carboxymethyl cellulase; gene; ds.
OS Bacillus subtilis.
OS Synthetic.
OS Chimeric.
XX Key Location/Qualifiers
FT promoter 1..460
FT /tag= a
FT /note= "CotG promoter"
FT CDS 461..2494
FT /tag= b
FT /partial
FT /product= "Bacillus subtilis spore coat protein cotG-
FT carboxymethyl cellulase (CMCCase) fusion protein"
FT /note= "This sequence lacks a start codon"
FT misc_feature 461..1045
FT /tag= c
FT /note= "CotG structural gene"
FT misc_feature 1046..1084
FT /tag= d
FT /note= "Linker"
FT misc_feature 1085..2491
FT /tag= e
FT /note= "CMCCase structural gene"
XX PN W0200246388-A1.

XX 13-JUN-2002.
XX 07-DEC-2001; 2001WO-KR002124.
XX 08-DEC-2000; 2000KR-00074835.
XX (GENO-) GENOFOCUS CO LTD.
XX Pan J, Choi S, Jung H;
XX WPI; 2002-463830/49.
XX P-PSDB; AAU98063.
XX Displaying a protein of interest on spore surface, by transforming a host
XX cell with a vector comprising a gene encoding a spore coat protein and a
XX gene encoding a protein of interest, useful for preparing protein
XX microarray.
XX Example 4; Page 145; 158pp; English.
XX The present invention relates to a new method for displaying a protein of
XX interest on a spore surface. The method of the invention involves
XX preparing a vector comprising a gene construct containing a gene encoding
XX a spore coat protein and a gene encoding a protein of interest,
XX transforming a host cell with the vector, displaying the protein of
XX interest on a surface of a spore of the host cell, and recovering the
XX spore. The method is useful for producing a delivery system for
XX recombinant live vaccine. The method is also useful for improving a
XX protein of interest using a resistance property of spore, in
XX bioconversion reactions, for preparing protein microarray, for producing
XX an antibody to antigen in vertebrates, and for preparing a whole cell
XX absorber. The spore coat proteins used circumvent a necessity for passage
XX across cell membrane, so that they do not need secretion signal and
XX target signal which are prerequisites of surface display motif, therefore
XX ensuring a surface display of protein, such as beta-galactosidase, in
XX orderly fashion, which is difficult to pass across cell membrane. The
XX present nucleic acid sequence encodes the Bacillus subtilis spore coat
XX protein cotG-carboxymethyl cellulase (CMCCase) fusion protein of the
XX invention
XX SQ Sequence 2510 BP; 869 A; 528 C; 521 G; 592 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 3.97e-13 Length: 2510
Score: 188.00 Matches: 34
Percent Similarity: 61.18% Conservatives: 18
Best Local Similarity: 40.00% Mismatches: 31
Query Match: 40.00% Indels: 2
DB: 6 Gaps: 1

US-09-917-376-4 (1-89) x ABK53202 (1-2510)
QY 4 GlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
DQ 2051 GGTATTCTGTACATAACAGCAGCGGATGGAGTATGACAGCAACCAATCCGTCGG 2110
QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
DQ 2111 CAGCTTCAATAAATAAATAACCGCAATACCGGTTGATTAAAGATGCTACTGCCGT 2170
QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaLa 63
DQ 2171 TACTGGTATACCGCAAAACAAAGCCCAAAACGTT-----GACTGTGACTACGGCGAG 2224
QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
DQ 2225 CTTGGATCGGCAATGTGACATACAAAGTTTGTGACGTTCATAAACCAAGCAAGTGCA 2284
QY 84 AspThrTyrLeuGln 88
DQ 2285 GATACCTATCTGGAA 2299
```

Search completed: May 12, 2004, 05:39:24  
Job time : 134.455 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 12, 2004, 04:33:52 ; Search time 835.94 Seconds  
(without alignments)  
3179.336 Million cell updates/sec

Title: US-09-917-376-4  
Perfect score: 470  
Sequence: 1 VSGGKVKQYKNDSAPGDNQ.....RASFGSVNPATPTADTVLQX 89

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame\_p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US09917376/runat\_11052004\_114535\_28249/app\_query.fasta\_1.1429  
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09917376@cgn\_1\_5334@runat\_11052004\_114535\_28249 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_est3:\*  
12: gb\_est4:\*  
13: gb\_est5:\*  
14: gb\_est6:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_nam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	81	17.2	584	14	CF480372	POL1_65_H
2	79.5	16.9	754	12	BQ047524	BQ047524 EST56642
3	78.5	16.7	768	14	CB293085	CB293085 UCRCS01_0
4	77	16.4	906	13	BQ722991	BQ722991 AGENCOURT
5	76	16.2	376	9	AA043716	AA043716 zk61f10.s
6	76	16.2	606	10	BE362891	BE362891 DGI_90_C1
7	76	16.2	641	14	CD226529	CD226529 CCI_46_E
8	76	16.2	657	10	BE362820	BE362820 DGI_89_C0
9	76	16.2	676	14	CF487649	CF487649 POL1_45_C
10	76	16.2	691	12	BG739392	BG739392 EM1_82_D0
11	76	16.2	724	28	BZ616323	BZ616323 iG58e08.g
12	76	16.2	814	12	BI218457	BI218457 602937951
13	76	16.2	905	28	BZ545807	BZ545807 OGALR72TM
14	76	16.2	934	29	CG270741	CG270741 OGMIB33TH
15	76	16.2	1001	29	CG270748	CG270748 OGMIB33TV
16	75.5	16.1	657	10	BF513128	BF513128 UI-H-BW1
17	75.5	16.1	1058	14	CK231562	CK231562 ILLUMIGEN
18	74.5	15.9	465	10	AW491464	AW491464 UI-M-BH3
19	74.5	15.9	613	14	CA595709	CA595709 wpaic.pko
20	74.5	15.9	650	28	BZ894843	BZ894843 Hg4_0131
21	74	15.7	302	14	CF076729	CF076729 QHK12L22
22	74	15.7	653	12	BI664037	BI664037 603289338
23	73.5	15.6	261	10	BF875908	BF875908 RC3-ET013
24	73.5	15.6	816	14	CB293980	CB293980 UCRCS01_0
25	73.5	15.6	819	14	CB293084	CB293084 UCRCS01_0
26	73.5	15.6	1039	12	BG334904	BG334904 602461219
27	73.5	15.6	1237	13	BQ065850	BQ065850 AGENCOURT
28	73	15.5	357	14	CF078232	CF078232 QHK1J23.Y
29	73	15.5	360	13	C64738	C64738 C64738 Yuj1
30	73	15.5	593	14	CF622108	CF622108 lafl3d06
31	73	15.5	776	29	BX209426	BX209426 Danilo rer
32	73	15.5	861	28	BH319803	BH319803 CH230-115
33	73	15.5	894	29	CG077853	CG077853 PUFKJ89TB
34	73	15.5	924	29	CG077856	CG077856 PUFKJ89TD
35	72.5	15.4	380	12	BG816375	BG816375 dad35808
36	72.5	15.4	450	9	AL799805	AL799805 AL799805
37	72.5	15.4	533	10	AW963120	AW963120 EST375193
38	72.5	15.4	544	9	AL898643	AL898643 AL898643
39	72.5	15.4	563	28	BZ424716	BZ424716 100018538
40	72.5	15.4	567	14	CF754312	CF754312 EST-77-2-
41	72.5	15.4	579	9	AL962576	AL962576 AL962576
42	72.5	15.4	601	29	CE253602	CE253602 tigr-gss-
43	72.5	15.4	640	9	AL846493	AL846493 AL846493
44	72.5	15.4	648	12	BM609423	BM609423 170006871
45	72.5	15.4	719	12	BG635313	BG635313 AT31857.5

ALIGNMENTS

RESULT 1  
CF480372  
LOCUS  
DEFINITION  
CF480372 POL1\_65\_H07.b1\_A002 Pollen Sorghum bicolor cDNA clone  
ACCESSION  
CF480372 POL1\_65\_H07\_A002 3', mRNA sequence.  
VERSION  
CF480372.1 GI:34509241  
KEYWORDS  
EST.  
SOURCE  
Sorghum bicolor (sorghum)  
ORGANISM  
Sorghum bicolor  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Sorghum.

Db	213	GCTTGTGTGAAGTGTGGTGTGCTTCTTTGGAGACGGTGTCTTGTCTGTGGTCCG	272
Qy	79	AlaThrProThrAlaAsp 84	
Db	273	AGCTTCAGCTCTGGTGAT 290	
RESULT 2			
LOCUS	BQ047524	754 bp	mRNA
DEFINITION	ESTS96642 P. infestans-challenged potato leaf, incompatible reaction Solanum tuberosum cDNA clone BPL117P18 5' end, mRNA sequence.		
ACCESSION	BQ047524		
VERSION	BQ047524.1	GI:19821510	
KEYWORDS	EST.		
SOURCE	Solanum tuberosum (potato)		
ORGANISM	Solanum tuberosum		
REFERENCE	1 (bases 1 to 754)		
AUTHORS	Zhang,P., Hernandez,M., Tornqvist,C.-E., Wirtz,U., Loukoianov,A., Rangel,P., Haberlach,G.T., Karamycheva,S.A., Tsai,J., Chiemingo,A., Bougri,O., Buell,C.R., Ronning,C.M., Helgeson,J. and Baker,B.		
TITLE	Generation of ESTs from Potato Leaves Challenged with Phytophthora infestans, incompatible interaction (2002)		
JOURNAL	Unpublished (2002)		
COMMENT	Contact: Robin Buell The Institute for Genomic Research 9712 Medical Center Dr, Rockville, MD 20850, USA Email: potato-array@tigr.org This clone can be obtained from the University of Arizona Genomics Institute. Orders can be made through URL: http://genome.arizona.edu/orders/ Seq primer: T3		
FEATURES	Location/Qualifiers		
source	1..754		
	/organism="Solanum tuberosum"		
	/mol_type="mRNA"		
	/cultivar="Kennebec"		
	/db_xref="taxon:4113"		
	/clone="BPL117P18"		
	/tissue_type="leaf"		
	/dev_stage="6 week old"		
	/lab_host="SOLR"		
	/clone_lib="P. infestans-challenged potato leaf, incompatible reaction"		
	/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; supplier: UC Berkeley, PGEC; sequencing: The Institute for Genomic Research. Whole plants were challenged with 450,000 sporangia/ml P. infestans isolate US-1 (US940501) in Biotron (Madison, Wisconsin). Leaf tissue was collected at 1, 2, 5, 12, and 24 hours post-challenge and frozen in liquid nitrogen immediately upon removal. Kennebec plants showed no signs of HR. Katahdin plants (susceptible to P. infestans US-1) were used as controls and showed infection. NOTE: We cannot exclude the possibility that this sequence is actually derived from Phytophthora rather than potato."		
ORIGIN			
Alignment Scores:			
Pred. No.:	27.9	Length:	754
Score:	79.50	Matches:	29
Percent Similarity:	38.05%	Conservative:	14
Best Local Similarity:	25.66%	Mismatches:	33
Query Match:	16.91%	Indels:	37
DB:	12	Gaps:	4
US-09-917-376-4 (1-89) x BQ047524 (1-754)			
Qy	7	ValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysProGlyLeuGln	26

```

FEATURES
source
Location/Qualifiers
1. .768
/organism="Citrus sinensis"
/mol_type="mRNA"
/cultivar="Washington navel"
/db_xref="taxon:2711"
/clone="UCRCS01.05cb12"
/tissue_type="Rind containing flavedo and albedo"
/dev_stage="Mature fruit"
/lab_host="E. coli TJC121"
/clone_lib="Washington Navel orange cold acclimated
flavedo & albedo cDNA library"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid;
Site.1: EcoRI; Site.2: XhoI; Plants were grown in the
field at University of California, Riverside Agricultural
Operations since 1983. The scion was Washington Navel
orange and the rootstock Carizzo Citrange. Tissue from
mature fruit was harvested at mid-day in January 2002
during a cold spell, when pre-dawn temperatures were
approximately -2 to 2 degree C. Approximately 2 cm median
sections of the rind were excised in the field from
several fruits, then wrapped in aluminum foil and frozen
quickly in dry ice. Total RNA was extracted using a phenol
extraction procedure described in J. Japanese Soc. Hort.
```

```

/clone="IMAGE:6191052"
/sex="male"
/tissue_type="sympathetic trunk"
/dev_stage="adult, 16 yr"
/lab_host="DH10B"
/clone_lib="Lupski_sympathetic_trunk"
/notes="Vector: PCW7-SPOF6 (Life Technologies); Site 1:
Not1; Site 2: SalI; cDNA made by oligo-dT priming.
Directionally cloned using the following adaptors:
5'-TCGACCCAGCGTCCG-3' and
5'-GACTAGTTCAGTCCGAGCGCGCCT(15)-3'. Size selected >
1 kb for average insert length 1.9 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine); available through Life
Technologies."

ORIGIN

Alignment Scores:
Pred. No.: 73.1 Length: 906
Score: 77.00 Matches: 16
Percent Similarity: 49.15% Conservative: 13
Best Local Similarity: 27.12% Mismatches: 26
Query Match: 16.38% Indels: 4
DB: 13 Gaps: 2

US-09-917-376-4 (1-89) x BQ722991 (1-906)

QY 7 ValGinTyrLysAsnAspSerAlaProGlyAspAsnGlnLeuGln 26
DB 548 CTGCCTATACCAATAAATCTCCAGGAGCGCCAGTGAGTA-----ATGAAG 598
QY 27 LeuValAsnThrGlySerSerValAspLeuSerThrValThrValArgTyrTrpPhe 46
DB 599 GCTGTAAAGTGGATGAAGAATATGATGGAAAGCTCTTCAGCACACTCTGTAC 658
QY 47 ThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAlaMetGly 65
DB 659 TCCCGG--GGCGGATGCGCCCGCAGCTCTACACCTGCATCTGGGGCAGTGTGGA 712

RESULT 5
AA043716 376 bp mRNA linear EST 11-MAY-1997
LOCUS zk61f10.s1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone
DEFINITION IMAGE:487339 3' similar to gb:L10240 BASIGIN PRECURSOR (HUMAN);,
mRNA sequence.
ACCESSION AA043716.1 GI:1521725
VERSION EST.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM

REFERENCE
AUTHORS Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 376)
Hillier,L., Lennon,G., Becker,M., Bonaldo,M.P., Chiapelli,B.,
Chisoso,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le.N.,
Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J.,
Trevisan,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R.
and Marra,M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
8889549
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 1400 Std Error: 0.00
Seq primer: -40M13 fwd. from Amersham
High quality sequence stop: 212.
Location/Qualifiers
1. .376
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:3761053"
/db_xref="taxon:9606"
/clone="IMAGE:487339"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="Soares_pregnant_uterus_NbHPU"
/notes="Organ: uterus; Vector: pT73-Pac; Site 1: Not 1;
Site 2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5',
AACTGGAAGAATTCCGGCGCGCTTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by M. Fatima Bonaldo."

ORIGIN

Alignment Scores:
Pred. No.: 28.5 Length: 376
Score: 76.00 Matches: 20
Percent Similarity: 43.42% Conservative: 13
Best Local Similarity: 26.32% Mismatches: 23
Query Match: 16.17% Indels: 20
DB: 9 Gaps: 3

US-09-917-376-4 (1-89) x AA043716 (1-376)

QY 17 GlyAspAsnGlnLeuLysProGlyLeuGlnLeuValAsnThrGlySerSerValAsp 36
DB 64 GGGGGTTCCACGTGAGCCCTGGT---GCTATAGANAGCCGCCCTCTCCAGGGGTGAG 120
QY 37 LeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySerSerThr 54
DB 121 CTGGGAGGGT---CTGGGGTCTGGAGTCCCGGCGATGGCGCAGTCTCCACGAAC 177
QY 55 -----LeuValTyrAsn 58
DB 178 CCCTCCAGAGTCCCGCGGATGCACAGACAGAGGGGGCTTGGGAGTGACTTGAGGC 237
QY 59 CysAspTrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGly 74
DB 238 TGTGACGGGTGCGCCCTCGGGTGTGGCAAGTGAAGTCTCTGTGGGC 285

RESULT 6
BE362891 606 bp mRNA linear EST 17-DEC-2001
LOCUS DGI_90_C11.g1_A002 Dark Grown 1 (DGI) Sorghum bicolor cDNA, mRNA
DEFINITION sequence.
ACCESSION BE362891
VERSION BE362891.2 GI:15724434
KEYWORDS EST.
SOURCE Sorghum bicolor (sorghum)
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 606)
Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and
Pratt,L.H.
An EST database from Sorghum: dark-grown seedlings
Unpublished (2000)
On Jul 20, 2000 this sequence version replaced gi:9304448.
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology

```

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 583 0210  
 Email: mmpratt@uga.edu  
 Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20. Three-prime sequences, which are obtained with PolyTMix or T7 sequencing primer, are presented as the reverse complement.  
 Seq primer: PolyTMix  
 High quality sequence start: 15  
 High quality sequence stop: 594  
 POLYA=No.

#### FEATURES

Location/Qualifiers  
 1..606  
 /organism="Sorghum bicolor"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:4558"  
 /clone\_lib="Dark Grown 1 (DGI)"  
 /note="Organ: 5-day-old dark-grown seedlings; Vector: Lambda Zap; Site 1: XhoI; Site 2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."

#### ORIGIN

Alignment Scores:  
 Pred. No.: 55.4 Length: 606  
 Score: 76.00 Matches: 23  
 Percent Similarity: 44.19% Conservative: 15  
 Best Local Similarity: 26.74% Mismatches: 38  
 Query Match: 16.17% Indels: 10  
 DB: 10 Gaps: 4

US-09-917-376-4 (1-89) x BE362891 (1-606)

QY 5 ValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnLeLysProGly 24  
 Db 208 ATAGCAATCAGCCGAACTCAATCAATCCGTCGGGAGGAGCCTGTCTGCGCCGGC 267  
 QY 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43  
 Db 268 -----CGCGCGGAGACAGAACCTTCCAAACGATCCGTCGAGGAA 315  
 QY 44 ---TyrTrp---PheThrArgAspGlyGlySerSerThrLeuValThrAsnCysAspTrp 61  
 Db 316 CCATGGTGGGATCTGCAGGAGGAGGTTTGACACATGGAATATATATATGATGA 375  
 QY 62 AlaAlaMetGlyCysGlyAsnLe-----ArgAlaSerPheGlySerValAsnPro 78  
 Db 376 GCTTGTGTGAAGTGTGTGACGTGGTCTTTGGAGACGGTGTCTGTCTGTGTCGCG 435  
 QY 79 AlaThrProThrAlaAsp 84  
 Db 436 AGCTTCAGTCTGGTGTAT 453

RESULT 7  
 LOCUS CD226529 641 bp mRNA linear EST 21-MAY-2003  
 DEFINITION CDNA clone CCCl\_46\_E07\_A007 Callus culture/cell suspension sorghum bicolor  
 ACCESSION CD226529  
 VERSION CD226529.1 GI:30969963  
 KEYWORDS EST.  
 SOURCE Sorghum bicolor (sorghum)  
 ORGANISM Sorghum bicolor  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum.  
 1 (bases 1 to 641)  
 Cordonnier-Pratt,M.-M., Wentzel,V., Suzuki,Y., Sugano,S., Klein,R.R., Liang,C., Sun,F., Sullivan,R., Shah,M., Rathore,K., Eastman,A. and Pratt,L.H.  
 An EST database from Sorghum: callus culture and cell suspension  
 Unpublished (2003)

#### COMMENT

Other\_ESTs: CCCl\_46\_E07.g1\_A007  
 Contact: Cordonnier-Pratt MM  
 Laboratory for Genomics and Bioinformatics  
 The University of Georgia, Department of Plant Biology  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 583 0210  
 Email: mmpratt@uga.edu  
 Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in the Human Genome Center, University of Tokyo Institute of Medical Science; plant material and RNA prepared at Texas A & M University; sequencing done in the laboratory for Genomics and Bioinformatics, University of Georgia. Sequence ends have been trimmed to exclude vector and regions below Phred quality 16. Three-prime sequences are presented as their reverse complement and have been trimmed to exclude polyA.  
 Seq primer: Sug3 (CGACCTGCAGCTCGAGCACA)  
 POLYA=Yes.

#### FEATURES

Location/Qualifiers  
 1..641  
 /organism="Sorghum bicolor"  
 /mol\_type="mRNA"  
 /cultivar="RTx430"  
 /db\_xref="taxon:4558"  
 /clone="CCCl\_46\_E07\_A007"  
 /lab\_host="DH10B-T1 phage-resistant E. coli"  
 /clone\_lib="Callus culture/cell suspension"  
 /note="Vector: pME18S-FL3; Site 1: XhoI; Site 2: XhoI; The library was prepared from a mixture of polyA+ RNA from callus culture tissue and cells in suspension culture. Double-stranded cDNA was cloned unidirectionally into different DraIII sites of the pME18S-FL3 vector (5-prime DraIII site is CACTGTGTG, 3-prime DraIII site is CACCATGTG). XhoI excises the cDNA insert."

#### ORIGIN

Alignment Scores:  
 Pred. No.: 59.9 Length: 641  
 Score: 76.00 Matches: 23  
 Percent Similarity: 44.19% Conservative: 15  
 Best Local Similarity: 26.74% Mismatches: 38  
 Query Match: 16.17% Indels: 10  
 DB: 14 Gaps: 4  
 US-09-917-376-4 (1-89) x CD226529 (1-641)

QY 5 ValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnLeLysProGly 24  
 Db 149 ATAGCAATCAGCCGAACTCAATCAATCCGTCGGGAGGAGCCTGTCTGCGCCGGC 208  
 QY 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43  
 Db 209 -----CGCGCGGAGACAGAACCTTCCAAACGATCCGTCGAGGAA 256  
 QY 44 ---TyrTrp---PheThrArgAspGlyGlySerSerThrLeuValThrAsnCysAspTrp 61  
 Db 257 CCATGGTGGGATCTGCAGGAGGAGGTTTGACACATGGAATATATATATATGATGA 316  
 QY 62 AlaAlaMetGlyCysGlyAsnLe-----ArgAlaSerPheGlySerValAsnPro 78  
 Db 317 GCTTGTGTGAAGTGTGTGACGTGGTCTTTGGAGACGGTGTCTGTCTGTGTCGCG 376  
 QY 79 AlaThrProThrAlaAsp 84  
 Db 377 AGCTTCAGTCTGGTGTAT 394  
 RESULT 8  
 LOCUS BE362820 657 bp mRNA linear EST 20-JUL-2000  
 DEFINITION DGI\_89\_C02.g2\_A002 Dark Grown 1 (DGI) Sorghum bicolor cDNA, mRNA sequence.  
 ACCESSION BE362820  
 VERSION BE362820.1 GI:9304377

```

KEYWORDS      EST.
SOURCE        Sorghum bicolor (sorghum)
ORGANISM      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
              clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE     1 (bases 1 to 657)
AUTHORS       Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and
              Pratt,L.H.
TITLE         An EST database from Sorghum: dark-grown seedlings
JOURNAL       Unpublished (2000)
COMMENT       Contact: Cordonnier-Pratt MM
              Laboratory for Genomics and Bioinformatics
              The University of Georgia, Department of Plant Biology
              Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
              Tel: 706 542 1860
              Fax: 706 583 0210
              Email: mmpratt@uga.edu
              Sequences have been trimmed to exclude PolyA, vector and regions
              below Phred quality 16. The threshold for highest quality sequence
              is 20.
              Seq primer: PolyTMix
              High quality sequence start: 89
              High quality sequence stop: 651
              POLYA=No.

FEATURES      Location/Qualifiers
              source
                1..657
                /organism="Sorghum bicolor"
                /mol_type="mRNA"
                /db_xref="taxon:4558"
                /clone_lib="Dark Grown 1 (DGI)"
                /note="Organ: 5-day-old dark-grown seedlings; Vector:
              Lambda Zap; Site 1: XhoI; Site 2: EcoRI; The library was
              made from poly-A RNA in the cloning vector lambda Zap II.
              Clones to be sequenced were prepared by mass excision."

ORIGIN
Alignment Scores:
Pred. No.:      62      Length:      657
Score:          76.00   Matches:     23
Percent Similarity: 44.19% Conservative: 15
Best Local Similarity: 26.74% Mismatches: 38
Query Match:    16.17% Indels:    10
DB:             10      Gaps:      4

US-09-917-376-4 (1-89) x BE362820 (1-657)

QY      5 VallysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysProGly 24
      ::: ::::: |||||
DB      204 ATAGCAATCAGCGCAAACTCAATGAATCCGGTCCGGAGAGGAGCGCTGTGCTGCCCGGCG 263
      ||| ::::: |||||
QY      25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg--- 43
      ||| ::::: |||||
DB      264 -----CGCGCCGGGACAGAACCTGAACTTCAACAGATCGGTGCAAGGAAA 311
      ||| ::::: |||||

QY      44 ---TyrTrp---PheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61
      ::::: |||||
DB      312 CCATGGTGGGATCTGCAGGGAAGAGGTTTGACGACATGGAATATATATATGATGA 371
      ||| ::::: |||||

QY      62 AlaAlaMetGlyCysGlyAsnIle-----ArgAlaSerPheGlySerValAsnPro 78
      ||| ::::: |||||
DB      372 GCTGTGTGAAGTGTGGTGACGTGTGTTCTTTTGAGACGCGTGTCTGTCTGTGTGTCG 431
      ||| ::::: |||||

QY      79 AlaThrProThrAlaAsp 84
      ::::: |||||
DB      432 AGCTTGACATCTGGTGAT 449
      ||| ::::: |||||

RESULT 9
CF487649
LOCUS
DEFINITION POLI_45_C08.b1_A002 Pollen Sorghum bicolor cDNA clone
              POLI_45_C08_A002 3', mRNA sequence.
ACCESSION CF487649

```

```

VERSION       CF487649.1 GI:34516518
KEYWORDS      EST.
SOURCE        Sorghum bicolor (sorghum)
ORGANISM      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
              clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE     1 (bases 1 to 676)
AUTHORS       Cordonnier-Pratt,M.-M., Suzuki,Y., Sugano,S., Klein,R.R., Liang,C.,
              Sun,F., Sullivan,R., Eastman,A., Cannon,R., Kern,B., Morgan,J.,
              Lucas,A., Al-Sheikh,A., Jones,V., Adibi,N., Owen,A., Gao,J. and
              Pratt,L.H.
TITLE         EST database from Sorghum: pollen
JOURNAL       Unpublished (2003)
COMMENT       Other ESTs: POLI_45_C08.g1_A002
              Contact: Cordonnier-Pratt MM
              Laboratory for Genomics and Bioinformatics
              The University of Georgia, Department of Plant Biology
              Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
              Tel: 706 542 1860
              Fax: 706 583 0210
              Email: mmpratt@uga.edu
              Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in
              the Human Genome Center, University of Tokyo Institute of Medical
              Science; plant material and RNA prepared at Texas A & M University;
              sequencing done in the Laboratory for Genomics and Bioinformatics,
              University of Georgia. Sequence ends have been trimmed to exclude
              vector and regions below Phred quality 16. Three-prime sequences
              are presented as their reverse complement and have been trimmed to
              exclude polyA.
              Seq primer: Sug3-14 (TAGTCTAGCGCGCGGACC)
              POLYA=Yes.

FEATURES      Location/Qualifiers
              source
                1..676
                /organism="Sorghum bicolor"
                /mol_type="mRNA"
                /cultivar="BTx623"
                /db_xref="taxon:4558"
                /clone="POLI_45_C08_A002"
                /lab_host="DH10B-T1 phage-resistant E. coli"
                /clone_lib="Pollen"
                /note="Organ: Pollen; Vector: pME18S-FL3; Site 1: XhoI;
              Site 2: XhoI; The library was prepared from polyA+ RNA
              from pollen at the late vacuolated-vacuolated stage of
              development. Pollen was harvested from greenhouse-grown
              panicles of sorghum line BTx623. Panicles were removed
              from the flag leaf prior to emergence, when no detectable
              amylase is present in pollen of male-fertile lines. This
              stage represents pollen collected from anthers about 8-14
              days prior to anthesis. Double-stranded cDNA was cloned
              unidirectionally into different DraIII sites of the
              pME18S-FL3 vector (5-prime DraIII site is CACTGTGTG,
              3-prime DraIII site is CACCATGTG). XhoI excises the cDNA
              insert."

ORIGIN
Alignment Scores:
Pred. No.:      64.6     Length:      676
Score:          76.00   Matches:     23
Percent Similarity: 44.19% Conservative: 15
Best Local Similarity: 26.74% Mismatches: 38
Query Match:    16.17% Indels:    10
DB:             14      Gaps:      4

US-09-917-376-4 (1-89) x CF487649 (1-676)

QY      5 VallysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysProGly 24
      ::: ::::: |||||
DB      214 ATAGCAATCAGCGCAAACTCAATGAATCCGGTCCGGAGAGGAGCGCTGTGCTGCCCGGCG 273
      ||| ::::: |||||

QY      25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg--- 43
      ||| ::::: |||||
DB      274 -----CGCGCCGGGACAGAACCTGAACTTCAACAGATCGGTGCAAGGAAA 321
      ||| ::::: |||||

```

325	-----CGCCCGGRCAGAACCTGNAATTCCTCAACAGATCGGTGCAAGAAA	372
Db		
44	---TyrTrp---PheThrArgAspGlySerSerThrLeuValTyrAsnCysAspTrp	61
Qy	:::	
Db		
373	CCATGCTGGGGGATCTGCAGGAAGGAGGTTTGAGCAACATGGAATATATATGATGAA	432
Qy	::	
Db		
62	AlaAlaMetGlyCysGlyAsnLeu-----ArgAlaSerPheGlySerValAsnPro	78
Qy	::	
Db		
433	GCTTGTGTGAAGTGTGTGACGTTGTTCTTTCGGACAGCGTGTCTTGCTCTGTGTGGTCCG	492
Qy		
79	AlaThrProThrAlaAsp	84
Qy	::: ::	
Db		
493	AGCTTGCAGTCTGGTGAT	510
Qy		

RESULT	11
BZ616323/c	
LOCUS	
DEFINITION	ZB616323 linear GSS 16-JAN-2003
	724 bp DNA
	1958e08.g1 WGS-ZmaySF (DHSA methyl filtered) Zea mays genomic clone
	1958e08 5', genomic survey sequence.
ACCESSION	ZB616323
VERSION	ZB616323.1 GI:27766878
KEYWORDS	GSS,
SOURCE	Zea mays
ORGANISM	Zea mays
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
	clade; Panicoideae; Andropogoneae; Zea.
REFERENCE	1 (bases 1 to 724)
AUTHORS	Rabinowicz,P.D., O'Shaughnessy,A.L., Ballija,V., Dedhia,N., Katzenburger,F., King,L., Miller,B., Muller,S., Nascimento,L., Zutavern,I., McCombie,W.R. and Marciniussen,R.A. Genomic shotgun sequences from Zea mays (methyl-filtered) Unpublished (2002)
TITLE	Contact: W. Richard McCombie
JOURNAL	
COMMENT	

Lita Annenberg Hazen Genome Sequencing Center  
Cold Spring Harbor Laboratory  
PO Box 100, Cold Spring Harbor, NY 11724, USA  
Tel: 516 367 8884  
Fax: 516 367 8874  
Email: mcombie@cshl.org  
Plate: i958 row: e column: 08  
Seq primer: -21M13univRev  
Class: shotgun  
High quality sequence stop: 728.

```

FEATURES
source
1. 724
Location/Qualifiers
/organism="Zea mays"
/mol_type="genomic DNA"
/cultivar="B73"
/db_xref="taxon:4577"
/clone="IG58e08"
/lab_host="DH5a"
/clone_lib="WGS-ZmaysF (DH5a methyl filtered)"
/note="Organ: immature ears; Site 1: Xba I; Site 2: Xba I;
The vector was digested with XbaI and one nucleotide was

```

added by fill in the recessive 3' end. The genomic DNA was nebulized, end repaired, adaptor ligated and size fractionated using sephadex. The resulting fragments were between 0.8 and 3 kb and were cloned into the vector (x/y reads in M13mp19, .b/g reads in pUC19). The same ligation was transformed into DH5a."

Pred. No.:	71	Length:	724
Score:	76.00	Matches:	34
Percent Similarity:	36.28%	Conservative:	7
Best Local Similarity:	30.09%	Mismatches:	29
Query Match:	16.17%	Indels:	43
DB:	28	Gaps:	6
US-09-917-376-4 (1-89) x B2616323 (1-724)			

```

QY 2 SerGlyGlyValLysValGlnTyrLysAsnAspSer-----AlaPro 16
Db ||||| ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
587 TCCGAGCGCTCAGATCCATACAGAGGAGGAGCAACGCTCGCGCAGCTGCGCC 528

QY 17 Gly-----AspAsnGlnIleLysProGlyLeuGlnLeuValAsn----- 29
Db ||||| ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
527 GCGGCCCGAGCTGAGCAGCTGCGAGGAGGAGCAGACTGCTTATCAACGCTGCGCCG 468

QY 30 Thr-GlySerSerValAsp----- 36
Db ||||| ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
467 ACGAGGAAGTTCGTCAACTCCCGGAGATCGTCGGAGCGGTCCAGAAGCTCGGTTCCGAG 408

QY 37 -----LeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySerSe 53
Db ||||| ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
407 GTCATCCCGATCGAGCGCGGTAGAGACCTTCACGCTGCGAGGACTTCGCCCGGACCGTCGAC 348

QY 53 rThrLeuValTyrAsnCysAspTrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPh 73
Db ||||| ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
347 TCGTGG-----ACGTGCTCATGGCGCGCGCGGG----- 317

QY 73 eGlySerValAsnProAlaThrProThrAlaAspThr 85
Db ||||| ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
316 -----CCGGCCTCACCAACTTCTTCTTCC 293

RESULT 12
BZ545807
LOCUS
DEFINITION BZ545807 814 bp mRNA linear EST 11-JUL-2001
602937951F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:5101069,
mRNA sequence.
ACCESSION BZ545807
VERSION BZ545807.1 GI:14671901
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 814)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11243 row: f column: 14
High quality sequence stop: 792.
FEATURES
source
1. .814
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5101069"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI CGAP Li9"
/note="Organ: liver; Vector: pCMV-Sport6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.9 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
ORIGIN
Alignment Scores:
Pred. No.: 83.6 Length: 814
Score: 76.00 Matches: 28
Percent Similarity: 44.93% Conservatives: 11
Best Local Similarity: 32.18% Mismatches: 29
Query Match: 16.17% Indels: 19
DB: 12 Gaps: 4

```

```

US-09-917-376-4 (1-89) x BZ545807 (1-814)
QY 2 SerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21
Db ||||| ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
33 GCAGGTGGAGCCCGAGTATCGGAGCGGACACACAGATCCCAAGGACATGACGAG--- 89

QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
Db ||||| ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
90 -----AGTTCCCGAGTGGACCTGAGGACCTG--- 116

QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeu-----ValTyrAsnCys 59
Db ||||| ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
117 CTCGGCTACTATACACACAGAGCGCGCGGCTCTCACAGATCCACGCGGATGTACGGCTGT 176

QY 60 Asp-TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAl 79
Db ||||| ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
177 GACGTGGGTGGACGCGCGC-----TCCTCCGGGTACGAGCAGGTGCGCA 224

QY 79 aThrProThrAlaAspThr 85
Db ||||| ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
225 TACGACGGCTGCGATTACA 243

RESULT 13
BZ545807
LOCUS
DEFINITION BZ545807 905 bp DNA linear GSS 16-DEC-2002
OCALR72TM ZM2.0.7.1.5 KB Zea mays genomic clone ZMMBMA0089L24,
genomic survey sequence.
ACCESSION BZ545807
VERSION BZ545807.1 GI:27099699
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 905)
REFERENCE
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
Consortium for Maize Genomics
Unpublished (2002)
Other GSSs: OGALR72TC
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR
Class: sheared ends.
FEATURES
source
1. .905
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBMA0089L24"
/clone_lib="ZM2.0.7.1.5 KB"
/note="Vector: pBSK-; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"
ORIGIN
Alignment Scores:
Pred. No.: 97 Length: 905
Score: 76.00 Matches: 34
Percent Similarity: 36.28% Conservatives: 7
Best Local Similarity: 30.09% Mismatches: 29
Query Match: 16.17% Indels: 43
DB: 28 Gaps: 6
US-09-917-376-4 (1-89) x BZ545807 (1-905)

```



QY 2 SerGlyGlyValIysValGlnTyrIysAsnAspSer-----AlaPro 16  
 Db 87 TCCGAGGCGTGCAGATCCCATACAGAGGAGCGCAACAGCAGCTGCGGAGCTGCGGCC 146  
 QY 17 Gly-----AspAsnGlnIleIysProGlyLeuGlnLeuValAsn----- 29  
 Db 147 GCGCCCCAGCTGAGCAGCTGCAGCGGAAGCCAGACTGATGCTTATCAACCGTGGCGCC 206  
 QY 30 Thr-GlySerSerValAsp----- 36  
 Db 207 ACGAGGAAGTTCGTCACTCCCGGAGATCGTCGGAGCGGTCAGAAAGCTGGGTTCGAG 266  
 QY 37 -----LeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySerSe 53  
 Db 267 GTCATCCGATCGAGCGCGTAGAGACCTCAGCGTGGAGGACTTCGCGCGGACCGTCGAC 326  
 QY 53 rThrLeuValTyrAsnCysAspTrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPh 73  
 Db 327 TCGTGGC-----ACGTGCTCATGGCGCGCACGGGG----- 357  
 QY 73 eGlySerValAsnProAlaThrProThrAlaAspThr 85  
 Db 358 -----CCGGCTCACCACACTTCTTCTTCC 381

RESULT 14  
 CG270741/c 934 bp DNA linear GSS 25-AUG-2003  
 LOCUS OGWIB33TH.ZM.0.7.1.5.KB.Zea.mays.genomic.clone.ZMMBMA0583E18,  
 DEFINITION genomic survey sequence.

ACCESSION CG270741  
 VERSION CG270741.1 GI:34182882  
 KEYWORDS GSS.

SOURCE Zea mays  
 ORGANISM Zea mays

REFERENCE  
 AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T., Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.  
 Consortium for Maize Genomics  
 Unpublished (2002)

TITLE JOURNAL  
 COMMENT Other GSSs: OGWIB33TH  
 Contact: Cathy Whitelaw  
 TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 301-838-5843  
 Fax: 301-838-0208  
 Email: whitelaw@tigr.org

Seq primer: TR  
 Class: sheared ends.

FEATURES  
 source Location/Qualifiers

1..934  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /strain="B73"  
 /db\_xref="taxon:4577"  
 /clone="ZMMBMA0583E18"  
 /clone\_lib="ZM.0.7.1.5.KB"  
 /note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb  
 methylation filtered genomic DNA library"

ORIGIN

Alignment Scores:  
 Pred. No.: 101 Length: 934  
 Score: 76.00 Matches: 34  
 Percent Similarity: 36.28% Conservative: 7  
 Best Local Similarity: 30.09% Mismatches: 29  
 Query Match: 16.17% Indels: 43  
 DB: 29 Gaps: 6

US-09-917-376-4 (1-89) x CG270741 (1-934)

QY 2 SerGlyGlyValIysValGlnTyrIysAsnAspSer-----AlaPro 16  
 Db 804 TCCGAGGCGTGCAGATCCCATACAGAGGAGCGCAACAGCAGCTGCGGAGCTGCGGCC 745  
 QY 17 Gly-----AspAsnGlnIleIysProGlyLeuGlnLeuValAsn----- 29  
 Db 744 GCGCCCCAGCTGAGCAGCTGCAGCGGAAGCCAGACTGATGCTTATCAACCGTGGCGCC 685  
 QY 30 Thr-GlySerSerValAsp----- 36  
 Db 684 ACGAGGAAGTTCGTCACTCCCGGAGATCGTCGGAGCGGTCAGAAAGCTGGGTTCGAG 625  
 QY 37 -----LeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySerSe 53  
 Db 624 GTCATCCGATCGAGCGCGTAGAGACCTCAGCGTGGAGGACTTCGCGCGGACCGTCGAC 565  
 QY 53 rThrLeuValTyrAsnCysAspTrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPh 73  
 Db 564 TCGTGGC-----ACGTGCTCATGGCGCGCACGGGG----- 534  
 QY 73 eGlySerValAsnProAlaThrProThrAlaAspThr 85  
 Db 533 -----CCGGCTCACCACACTTCTTCTTCC 510

RESULT 15

CG270748

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Other GSSs: OGWIB33TH

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TP

Class: sheared ends.

FEATURES

source Location/Qualifiers

1..1001

/organism="Zea mays"

/mol\_type="genomic DNA"

/strain="B73"

/db\_xref="taxon:4577"

/clone="ZMMBMA0583E18"

/clone\_lib="ZM.0.7.1.5.KB"

/note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb  
 methylation filtered genomic DNA library"

ORIGIN

Alignment Scores:

Pred. No.: 112 Length: 1001

Score: 76.00 Matches: 34

Percent Similarity: 36.28% Conservative: 7

Best Local Similarity: 30.09% Mismatches: 29

Query Match: 16.17% Indels: 43

DB: 29 Gaps: 6

US-09-917-376-4 (1-89) x CG270748 (1-1001)

```
QY 2 SerGlyValLysValGlnTyrLysAsnAspSer-----AlaPro 16
Db 418 TCCGAGCGGTCCAGATCCATACAAAGGAGGGAACAGCAACGCTGCGCAGCTGGCGCC 477
QY 17 Gly-----AspAsnGlnIleLysProGlyLeuGlnLeuValAsn----- 29
Db 478 GCGCGCCCGAGCTGAGCAGCTGCAGCGGGAAGCCACGACTGATGCTTATCAACCGTGCGCCG 537
QY 30 Thr-GlySerSerSerValAsp----- 36
Db 538 ACGAGGAAGTTCTGTTCAACTCCCGGAGATCGTCGGAGCGGTCCAGAAGGCTGGGTTTCGAG 597
QY 37 -----LeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySerSe 53
Db 598 GTCATCCCGATCGAGCGCGGTAGAGACCTCAGCGTGGAGACTTCGCCCGGACCGTCGAC 657
QY 53 rThrLeuValTyrAsnCysAspTrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPh 73
Db 658 TCGTGGC-----ACGTGCTCATGGCGCGCACGGGG----- 688
QY 73 eGlySerValAsnProAlaThrProThrAlaAspThr 85
Db 689 -----CCGGCCTCACCACACTCTCTTCTCC 712
```

Search completed: May 12, 2004, 11:39:00  
Job time : 840.94 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 12, 2004, 04:43:36 ; Search time 22.2257 Seconds  
(without alignments)  
2222.227 Million cell updates/sec

Title: US-09-917-376-4

Perfect score: 470

Sequence: 1 VSGGVKQYKXNDAPGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09917376/runat\_11052004\_114536\_28265/app\_query.fasta\_1.1429  
-DB=Issued Patents NA -QWMT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09917376@cgn\_1\_116@runat\_11052004\_114536\_28265 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:\*

1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PTUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	195.5	41.6	2029	3	US-09-136-574A-46
2	194.5	41.4	6416	3	US-09-136-574A-2
3	192.5	41.0	11707	3	US-09-136-574A-1
4	182.5	38.8	2977	2	US-07-862-588B-1
5	165	35.1	1438	4	US-09-339-159B-3
6	165	35.1	1482	3	US-09-198-956-9
7	165	35.1	1482	3	US-09-198-955A-11
8	165	35.1	1482	4	US-09-694-531-11
9	165	35.1	1482	4	US-09-670-141-9
10	165	35.1	1482	4	US-10-072-152-11
11	120.5	25.6	1624	2	US-07-862-588B-6
12	120.5	25.6	1775	2	US-07-862-588B-5

13	114	24.3	486	1	US-08-048-164A-1	Sequence 1, Appli
c 14	114	24.3	486	1	US-08-048-164A-3	Sequence 3, Appli
15	114	24.3	486	1	US-08-460-462-1	Sequence 1, Appli
c 16	114	24.3	486	1	US-08-460-462-3	Sequence 3, Appli
17	114	24.3	486	1	US-08-460-457-1	Sequence 1, Appli
c 18	114	24.3	486	1	US-08-460-457-3	Sequence 3, Appli
19	114	24.3	486	1	US-08-460-458-1	Sequence 1, Appli
c 20	114	24.3	486	1	US-08-460-458-3	Sequence 3, Appli
21	114	24.3	486	2	US-08-460-455-1	Sequence 1, Appli
c 22	114	24.3	486	2	US-08-460-455-3	Sequence 3, Appli
23	114	24.3	486	2	US-08-330-394A-1	Sequence 1, Appli
c 24	114	24.3	486	2	US-08-330-394A-3	Sequence 3, Appli
25	114	24.3	499	3	US-09-006-636-5	Sequence 5, Appli
26	114	24.3	499	3	US-09-006-632-5	Sequence 5, Appli
27	114	24.3	499	4	US-09-325-274-5	Sequence 21, Appl
28	113	24.0	1146	4	US-09-277-716-21	Sequence 21, Appl
29	113	24.0	1146	4	US-09-609-161B-21	Sequence 3, Appli
30	77	16.4	8107	4	US-09-335-586-3	Sequence 33, App
31	75.5	16.1	5714	4	US-09-620-312D-393	Sequence 21, Appl
32	72.5	15.4	20284	4	US-09-526-193A-21	Sequence 39, Appl
33	70	14.9	25165	4	US-09-453-702B-39	Sequence 20, Appl
34	67.5	14.4	2241	2	US-08-838-219B-20	Sequence 20, Appl
35	67.5	14.4	2241	3	US-09-233-336A-20	Sequence 20, Appl
36	67.5	14.4	2241	3	US-09-233-752A-20	Sequence 20, Appl
37	67.5	14.4	2241	3	US-09-402-036-20	Sequence 20, Appl
38	67.5	14.4	2241	4	US-09-904-226-20	Sequence 20, Appl
39	67.5	14.4	2370	2	US-08-838-219B-19	Sequence 19, Appl
40	67.5	14.4	2370	3	US-09-233-336A-19	Sequence 19, Appl
41	67.5	14.4	2370	3	US-09-233-752A-19	Sequence 19, Appl
42	67.5	14.4	2370	3	US-09-402-036-19	Sequence 19, Appl
43	67.5	14.4	2370	4	US-09-904-226-19	Sequence 19, Appl
44	67.5	14.4	118067	4	US-09-497-855A-32	Sequence 32, Appl
45	66	14.0	711	1	US-08-061-092A-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1  
US-09-136-574A-46  
; Sequence 46, Application US/09136574A  
; Patent No. 6294366  
; GENERAL INFORMATION:  
; APPLICANT: Farrington, Graham K.  
; Anderson, Paige  
; Gibbs, Moreland  
; Bergquist, Peter  
; Daniels, Roy  
; Morgan, Hugh W.  
; Williams, Diane P.  
; TITLE OF INVENTION: Compositions and Methods for  
; Treating Cellulose Containing Fabrics Using Truncated  
; Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49  
CORRESPONDENCE ADDRESS:  
ADDRESSES: Howson and Howson  
STREET: Spring House Corporate Center, P.O. Box 457  
CITY: Spring House  
STATE: PA  
COUNTRY: USA  
ZIP: 19477

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/136,574A  
FILING DATE: 19-Aug-1998  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/932,571  
FILING DATE: September 19, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Bak, Mary E.



```
; STATE: PA
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/136,574A
; FILING DATE: 19-Aug-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/932,571
; FILING DATE: September 19, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: 1997US001/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11707 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-136-574A-1

Alignment Scores:
Pred. No.: 3 15e-15 Length: 11707
Score: 192.50 Matches: 37
Percent Similarity: 62.35% Conservative: 16
Best Local Similarity: 43.53% Mismatches: 31
Query Match: 40.96% Indels: 1
DB: 3 Gaps: 1

US-09-917-376-4 (1-89) x US-09-136-574A-1 (1-11707)
QY 4 GlyVallyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
Db 4038 GGTGTTGAAGTACTATACAAAGCAATGAGCAAGTGCAGCAGCAAGTCTCTATAAGCGC 4097
QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
Db 4098 TGGTTTAAAGATAGTGAATGGAGGAGCAGCAGTGTGATCTTAGCAGGGTTAAGATAAGA 4157
QY 44 TyrTrpPheThrArgAspGlyGlySerThrLeuValTyrAsnCysAspTrpAla 63
Db 4158 TACTGTCACAGTGTGATGTCATCATTTTGAAGCTGAGCAGCGGAGTGTGAGCGC 4214
QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
Db 4215 ATAGGGGCAAGCAATGTGACATTCATTTTGAAGCTGAGCAGCGGAGTGTGAGCGC 4274
QY 84 AspThrTyrLeuGln 88
Db 4275 GATTATTACTTGGAG 4289

RESULT 4
US-07-862-588B-1
; Sequence 1, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linea
; APPLICANT: Sch lein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5916796o No. 5916796disk of No. 5916796th America, Inc.

; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
; REFERENCE/DOCKET NUMBER: 3425.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2977 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Bacillus lautus
; STRAIN: NCIMB 40250
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 677..2776
; OTHER INFORMATION:
; US-07-862-588B-1

Alignment Scores:
Pred. No.: 1 05e-14 Length: 2977
Score: 182.50 Matches: 36
Percent Similarity: 66.27% Conservative: 19
Best Local Similarity: 43.37% Mismatches: 25
Query Match: 38.83% Indels: 3
DB: 2 Gaps: 2

US-09-917-376-4 (1-89) x US-07-862-588B-1 (1-2977)
QY 6 LysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysProGlyLeu 25
Db 144 CGACTGCAGTACAGCGCGCGGATCAAAATGACGCGACCAACAGATCAAGCGCTCTTC 203
QY 26 GlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArgTyrTrp 45
Db 204 AACATCAAAAACAACCGTACTTCGGCTGTTGATTTAAGCAGCGCTCAAAATCCGCTACTAC 263
QY 46 PheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAlaMetGly 65
Db 264 TTCACCAAGATGGTTCTCGCGCGGTGAACGGCTGG---ATCGACTGGCGGACGCTCGGC 320
QY 66 CysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAlaAspThr 85
Db 321 GCAGCAACATTCAGATCTCGTTTGGC-----AACCTACTGCACGAATTCGATACG 374
QY 86 TyrLeuGln 88
Db 375 TACGTGGAG 383
```

SOFTWARE: FastSEQ for Windows Version 3.0

```
Alignment Scores:
Pred. No.: 9.37e-13 Length: 1482
Score: 165.00 Matches: 35
Percent Similarity: 59.79% Conservative: 23
Best Local Similarity: 36.08% Mismatches: 29
Query Match: 35.11% Indels: 10
DB: 3 Gaps: 2

US-09-917-376-4 (1-89) x US-09-198-955A-11 (1-1482)
QY 1 ValSerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 991 GTATCAGGCAATTGAAGGTTGAATTTCTACACAGCAATCCTTCAGATACTACTAATCA 1050
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1051 ATCAATCTCTCAGTTCAGGTTACTAATACCGGAGCAGTGCATTTGATTGTGCCAAACTC 1110
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1111 ACATTGAGATATTATTATACAGTACGCGACGAGAGATCAGACCTTCTGG---TGTGAC 1167
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1168 CATGCTGCAATAATCGGCAGTAACGCGAGTACAAACGGAATTACTTCAAAATGTAAGAAGGA 1227
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1228 ACATTGTGTAATAATGAGTTCTCTCAACAAATAACGCGACACCTTACCTTGAA 1278

RESULT 8
US-09-694-531-11
; Sequence 11, Application US/09694531
; Patent No. 6368843
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. 6368843el Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/09/694,531
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 09/198,955
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Clostridium thermocellum
US-09-694-531-11
Alignment Scores:
Pred. No.: 9.37e-13 Length: 1482
Score: 165.00 Matches: 35
Percent Similarity: 59.79% Conservative: 23
Best Local Similarity: 36.08% Mismatches: 29
Query Match: 35.11% Indels: 10
DB: 3 Gaps: 2

US-09-917-376-4 (1-89) x US-09-694-531-11 (1-1482)
QY 1 ValSerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 991 GTATCAGGCAATTGAAGGTTGAATTTCTACACAGCAATCCTTCAGATACTACTAATCA 1050
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1051 ATCAATCTCTCAGTTCAGGTTACTAATACCGGAGCAGTGCATTTGATTGTGCCAAACTC 1110
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1111 ACATTGAGATATTATTATACAGTACGCGACGAGAGATCAGACCTTCTGG---TGTGAC 1167
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1168 CATGCTGCAATAATCGGCAGTAACGCGAGTACAAACGGAATTACTTCAAAATGTAAGAAGGA 1227
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1228 ACATTGTGTAATAATGAGTTCTCTCAACAAATAACGCGACACCTTACCTTGAA 1278

RESULT 9
US-09-670-141-9
; Sequence 9, Application US/09670141
; Patent No. 6429000
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels Erik K.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Schnorr, Kirk
; TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus
; TITLE OF INVENTION: Licheniformis
; FILE REFERENCE: 5377.200-US
; CURRENT APPLICATION NUMBER: US/09/670,141
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: 09/198,956
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-670-141-9
Alignment Scores:
Pred. No.: 9.37e-13 Length: 1482
Score: 165.00 Matches: 35
Percent Similarity: 59.79% Conservative: 23
Best Local Similarity: 36.08% Mismatches: 29
Query Match: 35.11% Indels: 10
DB: 4 Gaps: 2

US-09-917-376-4 (1-89) x US-09-670-141-9 (1-1482)
QY 1 ValSerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 991 GTATCAGGCAATTGAAGGTTGAATTTCTACACAGCAATCCTTCAGATACTACTAATCA 1050
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1111 ACATTGAGATATTATTATACAGTACGCGACGAGAGATCAGACCTTCTGG---TGTGAC 1167
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1111 ACATTGAGATATTATTATACAGTACGCGACGAGAGATCAGACCTTCTGG---TGTGAC 1167
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1168 CATGCTGCAATAATCGGCAGTAACGCGAGTACAAACGGAATTACTTCAAAATGTAAGAAGGA 1227
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1228 ACATTGTGTAATAATGAGTTCTCTCAACAAATAACGCGACACCTTACCTTGAA 1278
```

Db 1051 ATCAATCTCAGTTCAAGTTACTAATACCGAAGCAGTGCAATTGATTGTCCAACTC 1110  
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
Db 1111 ACATTGAGATATTATTATACAGTAGACGACAGAAAGATCAGACCTTCTGG---TGTGAC 1167  
QY 61 TTPAlaAlaMet-----GlyCysGlyAsnIleArgAla 71  
Db 1168 CATGCTGCAATATCGGCAGTAAACGCGAGCTCAACAGGAATTACTTCAAAATGTAAAGGA 1227  
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88  
Db 1228 ACATTGTGTAATAATGAGTTCTTCAACAAATAACGCGACACCTACTCTTGAA 1278

## RESULT 10

US-10-072-152-11  
; Sequence 11, Application US/10072152  
; Patent No. 6677147  
; GENERAL INFORMATION:  
; APPLICANT: Andersen, Lene N.  
; APPLICANT: Schuelein, Martin  
; APPLICANT: Lange, Niels E.  
; APPLICANT: Bjornvad, Mads E.  
; APPLICANT: Moller, Soren  
; APPLICANT: Glad, Sanne O. S.  
; APPLICANT: Kauppinen, Markus S.  
; APPLICANT: Schnorr, Kirk  
; APPLICANT: Kongsbak, Lars  
; TITLE OF INVENTION: No. 6677147el Pectate Lyases  
; FILE REFERENCE: 5378 200-US  
; CURRENT APPLICATION NUMBER: US/10/072,152  
; CURRENT FILING DATE: 2002-02-07  
; PRIOR APPLICATION NUMBER: US/09/198,955  
; PRIOR FILING DATE: 1998-11-24  
; PRIOR APPLICATION NUMBER: 1343/97  
; PRIOR FILING DATE: 1997-11-24  
; PRIOR APPLICATION NUMBER: 1344/97  
; PRIOR FILING DATE: 1997-11-24  
; PRIOR APPLICATION NUMBER: 60/067,249  
; PRIOR FILING DATE: 1997-12-02  
; PRIOR APPLICATION NUMBER: 60/067,240  
; PRIOR FILING DATE: 1997-12-02  
; PRIOR APPLICATION NUMBER: 09/073,684  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR APPLICATION NUMBER: 09/184,217  
; PRIOR FILING DATE: 1998-11-02  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11  
; LENGTH: 1482  
; TYPE: DNA  
; ORGANISM: Clostridium thermocellum  
US-10-072-152-11

Alignment Scores:  
Pred. No.: 9,37e-13 Length: 1482  
Score: 165.00 Matches: 35  
Percent Similarity: 59.79% Conservative: 23  
Best Local Similarity: 36.08% Mismatches: 29  
Query Match: 35.11% Indels: 10  
DB: 4 Gaps: 2

US-09-917-376-4 (1-89) x US-10-072-152-11 (1-1482)

QY 1 ValSerGlyValValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20  
Db 991 GTATCAGGCAATTGGAAGTTGAATTTCTACACAGCAATCTTTCAGATACTACTACTCA 1050  
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40  
Db 1051 ATCAATCTCAGTTCAAGTTACTAATACCGAAGCAGTGCAATTGATTGTCCAAACTC 1110  
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60

Db 1111 ACATTGAGATATTATTATACAGTAGACGACAGAAAGATCAGACCTTCTGG---TGTGAC 1167  
QY 61 TTPAlaAlaMet-----GlyCysGlyAsnIleArgAla 71  
Db 1168 CATGCTGCAATATCGGCAGTAAACGCGAGCTCAACAGGAATTACTTCAAAATGTAAAGGA 1227  
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88  
Db 1228 ACATTGTGTAATAATGAGTTCTTCAACAAATAACGCGACACCTACTCTTGAA 1278

## RESULT 11

US-07-862-588B-6  
; Sequence 6, Application US/07862588B  
; Patent No. 5916796  
; GENERAL INFORMATION:  
; APPLICANT: Joergensen, Per Linnaa  
; APPLICANT: Sch lein, Martin  
; APPLICANT: Hansen, Christian  
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 59167960 No. 5916796disk of No. 5916796th America, Inc.  
; STREET: 405 Lexington Avenue, 62nd floor  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10017  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/862,588B  
; FILING DATE: 19920727  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DK 164/90  
; FILING DATE: 19-JAN-1990  
; PRIOR APPLICATION DATA: PCT/DK91/00013  
; FILING DATE: 18-JAN-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Zelson, Steve T. / Lambiris, Elias J.  
; REGISTRATION NUMBER: 30,335 / 33,728  
; REFERENCE/DOCKET NUMBER: 3425.204-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212 867 0123  
; TELEFAX: 212 867 0298  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1624 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: N-terminal  
; ORIGINAL SOURCE:  
; ORGANISM: Bacillus lautus  
; STRAIN: NCIMB 40250  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 30..1607  
; OTHER INFORMATION:  
US-07-862-588B-6

Alignment Scores:  
Pred. No.: 1.17e-06 Length: 1624  
Score: 120.50 Matches: 28  
Percent Similarity: 54.12% Conservative: 18  
Best Local Similarity: 32.94% Mismatches: 36  
Query Match: 25.64% Indels: 3



DB: 2 Gaps: 2

US-09-917-376--4 (1-89) x US-07-862-588B-6 (1-1624)

QY 2 SerGlyGlyValLysValGlnTyrLysAenAspSerAlaProGlyAspAenGinLe 21  
Db 1323 ACGGGGAACCTTGTTGCCAAATACAAGATTGGCGCAGACTAGCCGCACGGATAACCAANTG 1382  
QY 22 LysProGlyLeuGlnLeuValAenThrGlySerSerValAspLeuSerThrValThr 41  
Db 1383 AACCTTCCTTAACATCAAGAACAACGGGTACAAACCCCTGTTAACTGAGCGGCCTCAAG 1442  
QY 42 ValArgTyTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAenCysAspTrp 61  
Db 1443 CTTNNNNNNNNNNNNAAGAAC---GGACCTGCGGATATGAGCTCGATCGACTGG 1499  
QY 62 AlaAlaMetGlyCysGlyAsnIleAgaLasrPheGlySerValAenProAlaThrPro 81  
Db 1500 GCCCAATCGCGGAACGAATGTTCTGTCGGCATTC-----GCTAACTTTACCGGGAGT 1553  
QY 82 ThrAlaAspThrTyr 86  
Db 1554 AATACGGATACTTAC 1568

RESULT 12

US-07-862-588B-5  
; Sequence 5, Application US/07862588B  
; Patent No. 5916796  
; GENERAL INFORMATION:  
; APPLICANT: Joergensen, Per Linnae  
; APPLICANT: Sch lein, Martin  
; APPLICANT: Hansen, Christian  
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 59167960 No. 5916796disk of No. 5916796th America, Inc.  
; STREET: 405 Lexington Avenue, 62nd floor  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10017  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/862,588B  
; FILING DATE: 19920727  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DX 164/90  
; FILING DATE: 19-JAN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/DK91/00013  
; FILING DATE: 18-JAN-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Zelson, Steve T. / Lambiris, Elias J.  
; REGISTRATION NUMBER: 30,335 / 33,728  
; REFERENCE/DOCKET NUMBER: 3425.204-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212 867 0123  
; TELEFAX: 212 867 0298  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1775 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FRAGMENT TYPE: N-terminal  
; ORIGINAL SOURCE:  
; ORGANISM: Bacillus lautus

```

; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 486 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..486
; US-08-048-164A-1

Alignment Scores:
Pred. No.: 1.64e-06 Length: 486
Score: 114.00 Matches: 29
Percent Similarity: 52.58% Conservatives: 22
Best Local Similarity: 29.90% Mismatches: 34
Query Match: 24.26% Indels: 12
DB: 1 Gaps: 4

US-09-917-376-4 (1-89) x US-08-048-164A-1 (1-486)
QY 2 SerGlyValIysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21
Db 7 ACATCATCAATGTCAGTTGAATTTTACAACCTCTAACAATCAGCACAACAACTCAATT 66
QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
Db 67 ACACCAATTAATCAAAATTAATACTACACATCTGACAGTGATTTAAATTTAAATGACGTA 126
QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61
Db 127 GTTAGATATTATACACAGTGAATGTTGATACACAGGACAACTTCTGG---TGTGACCAT 183
QY 62 AlaAlaMetGlyCysGlyAsn-----IleArgAlaSerPhe 73
Db 184 GCTGTGTCATTATTAGGAATAGCTATGTTGATACACTAGCAAGTGACAGCAAACTTC 243
QY 74 -----GlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 244 GTTAAAGAAACAGCAAGCCCA---ACATCAACCTATGATACATATGTTGAA 291

RESULT 14
US-08-048-164A-1/c
; Sequence 3, Application US/08048164A
; Patent No. 5496934
; GENERAL INFORMATION:
; APPLICANT: Shoseyov, Oded
; APPLICANT: Shpiegl, Itai
; APPLICANT: Goldstein, Marc A.
; APPLICANT: Doi, Roy H.
; TITLE OF INVENTION: CELLULOSE BINDING DOMAIN
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/048,164A
; FILING DATE: 14-APR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:

```

```

; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7809-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 486 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; US-08-048-164A-3

Alignment Scores:
Pred. No.: 1.64e-06 Length: 486
Score: 114.00 Matches: 29
Percent Similarity: 52.58% Conservatives: 22
Best Local Similarity: 29.90% Mismatches: 34
Query Match: 24.26% Indels: 12
DB: 1 Gaps: 4

US-09-917-376-4 (1-89) x US-08-048-164A-3 (1-486)
QY 2 SerGlyValIysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21
Db 480 ACATCATCAATGTCAGTTGAATTTTACAACCTCTAACAATCAGCACAACAACTCAATT 421
QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
Db 420 ACACCAATTAATCAAAATTAATACTACACATCTGACAGTGATTTAAATTTAAATGACGTA 361
QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61
Db 360 GTTAGATATTATACACAGTGAATGTTGATACACAGGACAACTTCTGG---TGTGACCAT 304
QY 62 AlaAlaMetGlyCysGlyAsn-----IleArgAlaSerPhe 73
Db 303 GCTGTGTCATTATTAGGAATAGCTATGTTGATACACTAGCAAGTGACAGCAAACTTC 244
QY 74 -----GlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 243 GTTAAAGAAACAGCAAGCCCA---ACATCAACCTATGATACATATGTTGAA 196

RESULT 15
US-08-460-462-1
; Sequence 1, Application US/08460462
; Patent No. 5670623
; GENERAL INFORMATION:
; APPLICANT: Shoseyov, Oded
; APPLICANT: Shpiegl, Itai
; APPLICANT: Goldstein, Marc A.
; APPLICANT: Doi, Roy H.
; TITLE OF INVENTION: METHODS OF USE OF CELLULOSE BINDING DOMAIN PROTEINS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,462
; FILING DATE: concurrently herewith
; CLASSIFICATION: 435

```

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/048,164  
FILING DATE: 14-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7809-006  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 486 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..486  
US-08-460-462-1

Alignment Scores:  
Pred. No.: 1.64e-06 Length: 486  
Score: 114.00 Matches: 29  
Percent Similarity: 52.58% Conservative: 22  
Best Local Similarity: 29.90% Mismatches: 34  
Query Match: 24.28% Indels: 12  
DB: 1 Gaps: 4

US-09-917-376-4 (1-89) x US-08-460-462-1 (1-486)

QY	2	SerGlyGlyValLysValcLnTyrlYsAsnAspSerAlaProGlyAspAsnGlnIle	21
DB	7	ACATCATCAATGTCAGTTGAATTTTACAACTCTAACAAATCAGACAAACAACTCAATT	66
QY	22	LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr	41
DB	67	ACACCAATAATCAAAATTACTTAACACATCTGACGTGATTAAATTTAAATGACGTAAA	126
QY	42	ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp	61
DB	127	GTTAGATATTATTACACAAGTCATGGTACACAAGGACAAACTTTCTGG--TGTGACCAT	183
QY	62	AlaAlaMetGlyCysGlyAsn-----IleArgAlaSerPhe	73
DB	184	GCTGGTGCATTATTAGGAATAGCTATTGTTGATAACACTAGCAAAAGTGACAGCAAACTTC	243
QY	74	-----GlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln	88
DB	244	GTTAAGAAGAACAGCAAGCCCA---ACATCAACCTATGATACATATGTTGAA	291

Search completed: May 12, 2004, 15:15:25  
Job time : 27.2257 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 12, 2004, 09:26:52 ; Search time 136.654 Seconds  
(without alignments)  
2950.286 Million cell updates/sec

Title: US-09-917-376-4

Perfect score: 470

Sequence: 1 VSGGVKQVKNDSAPGDNQ.....RASFGSVNPTPTADTYLQX 89

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2941586 seqs, 2264995651 residues

Total number of hits satisfying chosen parameters: 5883172

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US09917376/runat\_11052004\_114537\_28336/app\_query.fasta\_1.1429  
-DB=Published Applications NA -QWMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09917376@cgn\_1\_809 @runat\_11052004\_114537\_28336  
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DRV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA.\*

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq2.\*  
14: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*  
17: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*  
18: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*  
19: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
------------	-------	-------------	-----------	----	-------------

RESULT 1  
US-09-917-378-2  
; Sequence 2, Application US/09917378  
; Publication NO. US20030119093A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: DECKER, STEPHEN R.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT MANNANASE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOLYTICUS  
; FILE REFERENCE: 40197.7US01  
; CURRENT APPLICATION NUMBER: US/09/917,378  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 2289  
; TYPE: DNA  
; ORGANISM: Acidothermus cellulolyticus  
US-09-917-378-2

ALIGNMENTS

1	469	99.8	2289	10	US-09-917-378-2	Sequence 2, Appli
2	466	99.1	3687	10	US-09-917-384-2	Sequence 2, Appli
3	466	99.1	3687	10	US-09-917-383-2	Sequence 2, Appli
4	462	98.3	2869	13	US-09-917-376-2	Sequence 2, Appli
5	462	98.3	2869	15	US-10-155-400-2	Sequence 2, Appli
6	257	54.7	2223	15	US-10-156-761-550	Sequence 550, App
7	257	54.7	9025608	15	US-10-156-761-1	Sequence 1, Appli
8	190	40.4	1527	16	US-10-369-493-46838	Sequence 46838, A
9	165	35.1	1438	13	US-10-372-054-3	Sequence 3, Appli
10	165	35.1	1482	13	US-10-655-433-11	Sequence 11, Appl
11	165	35.1	1482	14	US-10-072-152-11	Sequence 11, Appl
12	157	33.4	1314	9	US-09-974-300-684	Sequence 684, App
13	155.5	33.1	4059	10	US-09-784-554B-3	Sequence 1, Appli
14	150.5	32.0	4056	10	US-09-784-554B-3	Sequence 3, Appli
15	130.5	27.8	4452	13	US-10-282-122A-15299	Sequence 15299, A
16	114	24.3	768	15	US-10-261-446-5	Sequence 5, Appli
17	113	24.0	1146	10	US-09-808-898-21	Sequence 21, Appl
18	78.5	16.7	3147	13	US-10-282-122A-13716	Sequence 13716, A
19	76	16.2	96597	12	US-10-052-483-226	Sequence 226, App
20	75.5	16.1	1041	16	US-10-260-238-742	Sequence 742, App
21	75.5	16.1	2091	13	US-10-211-462-222	Sequence 222, App
22	75.5	16.1	3396	17	US-10-398-458-1	Sequence 1, Appli
23	75.5	16.1	4041	9	US-09-747-835A-21	Sequence 21, Appl
24	75.5	16.1	4041	13	US-10-312-312-21	Sequence 21, Appl
25	75.5	16.1	4353	15	US-10-225-567A-405	Sequence 405, App
26	75.5	16.1	4914	9	US-09-974-298-188	Sequence 188, App
27	75.5	16.1	4914	13	US-10-116-802-38	Sequence 38, Appl
28	75.5	16.1	5714	9	US-09-747-835A-19	Sequence 19, Appl
29	75.5	16.1	5714	9	US-09-747-835A-60	Sequence 60, Appl
30	75.5	16.1	5714	13	US-10-312-312-19	Sequence 19, Appl
31	75.5	16.1	5714	13	US-10-312-312-60	Sequence 60, Appl
32	75.5	16.1	5714	15	US-10-037-270-393	Sequence 393, App
33	75.5	16.1	5714	16	US-10-117-722-393	Sequence 393, App
34	75.5	16.1	5773	9	US-09-747-835A-18	Sequence 18, Appl
35	75.5	16.1	5773	13	US-10-312-312-18	Sequence 18, Appl
36	75	16.0	484	13	US-10-424-599-54043	Sequence 54043, A
37	73	15.5	9025608	15	US-10-156-761-1	Sequence 1, Appli
38	72.5	15.4	402	9	US-09-764-860-85	Sequence 85, Appl
39	72.5	15.4	402	15	US-10-074-095-85	Sequence 85, Appl
40	72.5	15.4	402	16	US-10-212-872-85	Sequence 85, Appl
41	72.5	15.4	9181	10	US-09-984-827-10	Sequence 10, Appl
42	72.5	15.4	20284	13	US-10-617-334-21	Sequence 21, Appl
43	72.5	15.4	20284	16	US-10-452-510-21	Sequence 21, Appl
44	72	15.3	2402	13	US-10-425-114-9462	Sequence 9462, Ap
45	72	15.3	48667	9	US-09-822-268A-3	Sequence 3, Appli

## Alignment Scores:

Pred. No.: 2,21e-56 Length: 2289  
Score: 469.00 Matches: 88  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 99.79% Indels: 0  
DB: 10 Gaps: 0

US-09-917-376-4 (1-89) x US-09-917-378-2 (1-2289)

QY 1 ValSerGlyGlyVallyValGlnTyrLysAsnAsnAspSerAlaProGlyAspAenGln 20  
DB 1363 GTGTGGGTGGGTGAAGTGCAGTACAGAACATGATTCGGCGCGGGTGATAACCG 1422  
QY 21 IleLysProGlyLeuGlnLeuValAenThrGlySerSerValAspLeuSerThrVal 40  
DB 1423 ATCAAAACCGGGTCTCCAGTTGGTGAATACGGGGTGTCTCGGTGGATTGTTCGACGGTG 1482  
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
DB 1483 ACGGTGCGGTACTGGTTTCCCGGGATGCGGTGCGTGCACACTGGTGTAACAACCTGTGAC 1542  
QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80  
DB 1543 TGGCGCGCGATGGGGTGGGAATATCCGGCTCGTTCGGTGAACCCGCGGACG 1602  
QY 81 ProThrAlaAspThrTyrLeuGln 88  
DB 1603 CCGACGGCGGACACCTACCTGCAG 1626

## RESULT 2

US-09-917-384-2

; Sequence 2, Application US/09917384  
; Publication No. US20030096342A1

## GENERAL INFORMATION:

; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: DECKER, STEPHEN R.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOXYTICUS  
; FILE REFERENCE: 40170.6US01  
; CURRENT APPLICATION NUMBER: US/09/917,384  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 3687  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Segment of  
US-09-917-384-2

## Alignment Scores:

Pred. No.: 1.13e-55 Length: 3687  
Score: 466.00 Matches: 87  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 98.86% Mismatches: 0  
Query Match: 99.15% Indels: 0  
DB: 10 Gaps: 0

US-09-917-376-4 (1-89) x US-09-917-384-2 (1-3687)

QY 1 ValSerGlyGlyVallyValGlnTyrLysAsnAsnAspSerAlaProGlyAspAenGln 20  
DB 1750 GTGTGGGTGGGTGAAGTGCAGTACAGAACATGATTCGGCGCGGGTGATAACCG 1809  
QY 21 IleLysProGlyLeuGlnLeuValAenThrGlySerSerValAspLeuSerThrVal 40  
DB 1810 ATCAAAACCGGGTCTCCAGTTGGTGAATACCGGGTGTCTCGTTCGGTGGATTGTTCGACGGTG 1869

QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
DB 1870 ACGGTGCGGTACTGGTTTCCCGGGATGCGGTGCGTGCACACTGGTGTCAACTGTGAC 1929  
QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80  
DB 1930 TGGCGCGCGATGGGGTGGGAATATCCGGCTCGTTCGGTGAACCCGCGGACG 1989  
QY 81 ProThrAlaAspThrTyrLeuGln 88  
DB 1990 CCGACGGCGGACACCTACCTGCAG 2013

## RESULT 3

US-09-917-383-2

; Sequence 2, Application US/09917383  
; Publication No. US20030104522A1

## GENERAL INFORMATION:

; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: DECKER, STEPHEN R.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOXYTICUS  
; FILE REFERENCE: 40170.6US01  
; CURRENT APPLICATION NUMBER: US/09/917,383  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 3687  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Segment of  
US-09-917-383-2

## Alignment Scores:

Pred. No.: 1.13e-55 Length: 3687  
Score: 466.00 Matches: 87  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 98.86% Mismatches: 0  
Query Match: 99.15% Indels: 0  
DB: 10 Gaps: 0

US-09-917-376-4 (1-89) x US-09-917-383-2 (1-3687)

QY 1 ValSerGlyGlyVallyValGlnTyrLysAsnAsnAspSerAlaProGlyAspAenGln 20  
DB 1750 GTGTGGGTGGGTGAAGTGCAGTACAGAACATGATTCGGCGCGGGTGATAACCG 1809  
QY 21 IleLysProGlyLeuGlnLeuValAenThrGlySerSerValAspLeuSerThrVal 40  
DB 1810 ATCAAAACCGGGTCTCCAGTTGGTGAATACCGGGTGTCTCGTTCGGTGGATTGTTCGACGGTG 1869  
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
DB 1870 ACGGTGCGGTACTGGTTTCCCGGGATGCGGTGCGTGCACACTGGTGTCAACTGTGAC 1929  
QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80  
DB 1930 TGGCGCGCGATGGGGTGGGAATATCCGGCTCGTTCGGTGAACCCGCGGACG 1989  
QY 81 ProThrAlaAspThrTyrLeuGln 88  
DB 1990 CCGACGGCGGACACCTACCTGCAG 2013

## RESULT 4

US-09-917-376-2

; Sequence 2, Application US/09917376  
; Publication No. US20040038334A1

```
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: 40197.4US01
; CURRENT APPLICATION NUMBER: US/09/917,376
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2869
; TYPE: DNA
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (2869)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-09-917-376-2

Alignment Scores:
Pred. No.: 2,98e-55 Length: 2869
Score: 462.00 Matches: 86
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 97.73% Mismatches: 0
Query Match: 98.30% Indels: 0
DB: 15 Gaps: 0

US-09-917-376-4 (1-89) x US-09-917-376-2 (1-2869)
QY 1 ValSerGlyGlyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 2605 GTGTCGGGTGGGGTGAAGTGCAGTATAGAATAATGATTCGGCGCGGGTGATATACAG 2664

QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 2665 ATCAAGCGGGTTTGCAGGTGGTGAATACCGGGTCGTCGTCGGTGGATTGTCGACGGTG 2724

QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 2725 ACGGTGCGGTACTGTTTCACCGGGATGGTGGCTCGTCGACACTGGTGTACACTGTGAC 2784

QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 2785 TGGCGGGCGATCGGGTGGGAATATCCGGCCCTCGTTCGGCTCGGTGAACCCGGCGACG 2844

QY 81 ProThrAlaAspThrTyrLeuGln 88
Db 2845 CCGACGGCGGACACCTACTCTGCAG 2868

RESULT 5
US-10-155-400-2
; Sequence 2, Application US/10155400
; Publication No. US20030108988A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: NREL 01-36A
; CURRENT APPLICATION NUMBER: US/10/155,400
; CURRENT FILING DATE: 2002-10-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2869
; TYPE: DNA
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (2869)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-09-917-376-2

Alignment Scores:
Pred. No.: 2,98e-55 Length: 2869
Score: 462.00 Matches: 86
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 97.73% Mismatches: 0
Query Match: 98.30% Indels: 0
DB: 15 Gaps: 0

US-09-917-376-4 (1-89) x US-09-917-376-2 (1-2869)
QY 1 ValSerGlyGlyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 2605 GTGTCGGGTGGGGTGAAGTGCAGTATAGAATAATGATTCGGCGCGGGTGATATACAG 2664

QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 2665 ATCAAGCGGGTTTGCAGGTGGTGAATACCGGGTCGTCGTCGGTGGATTGTCGACGGTG 2724

QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 2725 ACGGTGCGGTACTGTTTCACCGGGATGGTGGCTCGTCGACACTGGTGTACACTGTGAC 2784

QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 2785 TGGCGGGCGATCGGGTGGGAATATCCGGCCCTCGTTCGGCTCGGTGAACCCGGCGACG 2844

QY 81 ProThrAlaAspThrTyrLeuGln 88
Db 2845 CCGACGGCGGACACCTACTCTGCAG 2868

RESULT 6
US-10-156-761-550
; Sequence 550, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 550
; LENGTH: 2223
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2223)
US-10-156-761-550

Alignment Scores:
Pred. No.: 3,69e-26 Length: 2223
Score: 257.00 Matches: 49
Percent Similarity: 71.26% Conservative: 13
Best Local Similarity: 56.32% Mismatches: 25
Query Match: 54.68% Indels: 0
DB: 15 Gaps: 0
```

```
US-09-917-376-4 (1-89) x US-10-156-761-550 (1-2223)
QY 2 SerGlyValValysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21
Db 1771 TCGGGCGGCTCAAGTCTCTACAAGAACACAGCTCTCGGCCACCGACGCGCATC 1830
QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
Db 1831 CGGCCAGGCTTCGGATCGTCAACCGCGAGCGGCTCCCTCGACCTGTCCAAAGGTACG 1890
QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTyr 61
Db 1891 GCCCGCTACTACTTACCCGGGACAGCGGCTCCGCCACCGTGAACCGCTGTGCGACTAC 1950
QY 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81
Db 1951 GCGGCGTGGCTGTTCACAGCTGAGCGGCTCCGCCACCGTGAACCGCTGTGCGACTAC 2010
QY 82 ThrAlaAspThrTyrLeuGln 88
Db 2011 GGAGCGGACGCTTACCTCGAA 2031

RESULT 7
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Alignment Scores:
Pred. No.: 2,72e-21 Length: 9025608
Score: 257.00 Matches: 49
Percent Similarity: 71.26% Conservative: 13
Best Local Similarity: 56.32% Mismatches: 25
Query Match: 54.68% Indels: 0
Gaps: 0
DB: 15

US-09-917-376-4 (1-89) x US-10-156-761-1 (1-9025608)
QY 2 SerGlyValValysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21
Db 706216 TCGGGCGGCTCAAGTCTCTACAAGAACACAGCTCTCGGCCACCGACGCGCATC 706275
QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
Db 706276 CGGCCAGGCTTCGGATCGTCAACCGCGAGCGGCTCCCTCGACCTGTCCAAAGGTACG 706335
QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTyr 61
Db 706336 GCGCGCTACTACTTACCCGGGACAGCGGCTCCGCCACCGTGAACCGCTGTGCGACTAC 706395

QY 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81
Db 706396 GCGGCGTGGCTGTTCACAGCTGAGCGGCTCCGCCACCGTGAACCGCTGTGCGCG 706455
QY 82 ThrAlaAspThrTyrLeuGln 88
Db 706456 GGAGCGGACGCTTACCTCGAA 706476

RESULT 8
US-10-369-493-46838
; Sequence 46838, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 46838
; LENGTH: 1527
; TYPE: DNA
; ORGANISM: Bacillus subtilis
US-10-369-493-46838

Alignment Scores:
Pred. No.: 8.02e-17 Length: 1527
Score: 190.00 Matches: 35
Percent Similarity: 65.12% Conservative: 21
Best Local Similarity: 40.70% Mismatches: 26
Query Match: 40.43% Indels: 4
Gaps: 2
DB: 16

US-09-917-376-4 (1-89) x US-10-369-493-46838 (1-1527)
QY 4 GlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
Db 1084 GGTATTCTGTACGTACAGAGGAGGATGATGACAGACCAACCAATCGTCCG 1143
QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
Db 1144 CAGCTTCAATAAAAAATAACGGCAATACCAACGGTTGATTTAAAAAGATGTCACTGCCGT 1203
QY 44 TyrTrpPhe---ThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAla 62
Db 1204 TACTGGTATAAGCAAAAAACAAAGCCCAAAAC-----TTTGACTGTGACTACGCG 1254
QY 63 AlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThr 82
Db 1255 CAGATTGGATCGGCAATGTGACACACAAGTTTGTGACGTTCATTAACCAACGAAGT 1314
QY 83 AlaAspThrTyrLeuGln 88
Db 1315 GCAGATACCTATCTGGAA 1332

RESULT 9
US-10-372-054-3
; Sequence 3, Application US/10372054
; Publication No. US20030203466A1
; GENERAL INFORMATION:
; APPLICANT: Kauppinen, Markus
; APPLICANT: Schulein, Martin
; APPLICANT: Schnorr, Kirk
; APPLICANT: Andersen, Lene
```

```
; APPLICANT: Bjornvad, Mads
; TITLE OF INVENTION: Novel Mannanases
; FILE REFERENCE: 5440.204-US
; CURRENT APPLICATION NUMBER: US/10/372,054
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: US/09/339,159B
; PRIOR FILING DATE: 1999-06-24
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 1438
; TYPE: DNA
; ORGANISM: Bacillus sp. I633
; US-10-372-054-3

Alignment Scores:
Pred. No.: 2,72e-13 Length: 1438
Score: 165.00 Matches: 35
Percent Similarity: 59.79% Conservative: 23
Best Local Similarity: 36.08% Mismatches: 29
Query Match: 35.11% Indels: 10
DB: 13 Gaps: 2

US-09-917-376-4 (1-89) x US-10-372-054-3 (1-1438)
QY 1 ValSerGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 940 GTATCAGCAATTGAGGTGAATCTACACAGCAATCCTTCAGATACTACTACTCA 999
QY 21 IleLysProGlyLeuGlnValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1000 ATCAATCCTCAGTCAAGTTACTAATACCGAAGCAGTGAATGATTGTCCAAACTC 1059
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1060 ACATTGAGATATTATTATACAGTAGCAGCAGAGAAGATCAGACCTTCTGG---TGTGAC 1116
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1117 CATGCTGCATATATCGGCAGTACCGCAGCTACACCGGAATTAATCAATGTAAAGGA 1176
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1177 ACATTGTAAATAAGTTCTCTCAACAATAACGACGACACCTTACCTTGAA 1227

RESULT 10
US-10-655-433-11
; Sequence 11, Application US/10655433
; Publication No. US2004006752A1
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schuelein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. US2004006752A1e1 Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/10/655,433
; CURRENT FILING DATE: 2003-09-04
; PRIOR APPLICATION NUMBER: US/09/198,955A
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02

US-09-917-376-4 (1-89) x US-10-655-433-11 (1-1482)
QY 1 ValSerGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 991 GTATCAGCAATTGAGGTGAATCTACACAGCAATCCTTCAGATACTACTACTCA 1050
QY 21 IleLysProGlyLeuGlnValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1051 ATCAATCCTCAGTCAAGTTACTAATACCGAAGCAGTGAATGATTGTCCAAACTC 1110
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1111 ACATTGAGATATTATTATACAGTAGCAGCAGAGAAGATCAGACCTTCTGG---TGTGAC 1167
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1168 CATGCTGCATATATCGGCAGTACCGCAGCTACACCGGAATTAATCAATGTAAAGGA 1227
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1228 ACATTGTAAATAAGTTCTCTCAACAATAACGACGACACCTTACCTTGAA 1278

RESULT 11
US-10-072-152-11
; Sequence 11, Application US/10072152
; Publication No. US20020142438A1
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schuelein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. US20020142438A1e1 Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/10/072,152
; CURRENT FILING DATE: 2002-02-07
; PRIOR APPLICATION NUMBER: US/09/198,955
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
```



```
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Clostridium thermocellum
US-10-072-152-11

Alignment Scores:
Pred. No.: 2,848-13 Length: 1482
Score: 165.00 Matches: 35
Percent Similarity: 59.79% Conservative: 23
Best Local Similarity: 36.08% Mismatches: 29
Query Match: 35.11% Indels: 10
DB: 14 Gaps: 2

US-09-917-376-4 (1-89) x US-10-072-152-11 (1-1482)

QY 1 ValSerGlyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 991 GTATCAGGCAATTTGAAGTTCAATCTCAACAGCAATCCTTCAGATACTACTCA 1050
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1051 ATCAATCTCAGTCAAGGTTTACTAATACCGGAAGCAGTGCAATTTGTTCCAAACTC 1110
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1111 ACATTGAGATATATATATACAGTAGCAGCAGAAAGATCAGACCTTCCTGG---TGTGAC 1167
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1168 CATGCTGCAATAATCGGCAGTAACGCGCAGCTACACAGGAATTACTTCAAAATGTAAAGGA 1227
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1228 ACATTGTAAATAATGAGTTCTCTCAACAAATAACGCAGACCTACCTTGAA 1278

RESULT 12
US-09-974-300-684
; Sequence 684, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: Expression
; FILE REFERENCE: 10085-500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 684
; LENGTH: 1314
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-684

Alignment Scores:
Pred. No.: 3,348-12 Length: 1314
Score: 157.00 Matches: 32
Percent Similarity: 56.47% Conservative: 16
Best Local Similarity: 37.65% Mismatches: 33
Query Match: 33.40% Indels: 4
DB: 9 Gaps: 2

US-09-917-376-4 (1-89) x US-09-974-300-684 (1-1314)

; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Clostridium thermocellum
US-10-072-152-11

Alignment Scores:
Pred. No.: 2,848-13 Length: 1482
Score: 165.00 Matches: 35
Percent Similarity: 59.79% Conservative: 23
Best Local Similarity: 36.08% Mismatches: 29
Query Match: 35.11% Indels: 10
DB: 14 Gaps: 2

US-09-917-376-4 (1-89) x US-10-072-152-11 (1-1482)

QY 1 ValSerGlyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 991 GTATCAGGCAATTTGAAGTTCAATCTCAACAGCAATCCTTCAGATACTACTCA 1050
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1051 ATCAATCTCAGTCAAGGTTTACTAATACCGGAAGCAGTGCAATTTGTTCCAAACTC 1110
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1111 ACATTGAGATATATATATACAGTAGCAGCAGAAAGATCAGACCTTCCTGG---TGTGAC 1167
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1168 CATGCTGCAATAATCGGCAGTAACGCGCAGCTACACAGGAATTACTTCAAAATGTAAAGGA 1227
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1228 ACATTGTAAATAATGAGTTCTCTCAACAAATAACGCAGACCTACCTTGAA 1278

RESULT 13
US-09-784-554B-1
; Sequence 1, Application US/09784554B
; Publication No. US20030032162A1
; GENERAL INFORMATION:
; APPLICANT: Schnorr, Kirk
; APPLICANT: Jorgensen, Per Lina
; APPLICANT: Schulein, Martin
; TITLE OF INVENTION: FAMILY 44 XYOLOGUCANASES
; FILE REFERENCE: 10017.200-US
; CURRENT APPLICATION NUMBER: US/09/784,554B
; CURRENT FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 4059
; TYPE: DNA
; ORGANISM: Paenibacillus polymyxa
US-09-784-554B-1

Alignment Scores:
Pred. No.: 2,5e-11 Length: 4059
Score: 155.50 Matches: 27
Percent Similarity: 56.47% Conservative: 21
Best Local Similarity: 31.76% Mismatches: 36
Query Match: 33.09% Indels: 1
DB: 10 Gaps: 1

US-09-917-376-4 (1-89) x US-09-784-554B-1 (1-4059)

QY 4 GlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
Db 3598 GGACTGCTGCTCAATATCGCAGCAGATACTAAGGTGAACGATATATCCTCAATCCG 3657
QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValArg 43
Db 3658 CAATTCCAAATTTGTAAACAAAGGCACACCTCCATCCGATCAGATCAACAGTTGAAATTCG 3717
QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAla 63
Db 3718 TACTACTACAAATCGACGGTACCGGAGCAG---ACTTTCACACTGCGACTATCGAGC 3774
QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
Db 3775 CTGAGCTGCTCAAGCTGAACGGTAAACTGTTAAATGGAGAGCGCTGCAACCGGTGCC 3834
QY 84 AspThrTyrLeuGln 88
Db 3835 GATTATTATTGGAA 3849

RESULT 14
US-09-784-554B-3
; Sequence 3, Application US/09784554B
```

```
; Publication No. US20030032162A1
; GENERAL INFORMATION:
; APPLICANT: Schnorr, Kirk
; APPLICANT: Jorgensen, Per Lina
; APPLICANT: Schulein, Martin
; TITLE OF INVENTION: FAMILY 44 XYLOGLUCANASES
; FILE REFERENCE: 10017.200-US
; CURRENT APPLICATION NUMBER: US/09/784,554B
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 4056
; TYPE: DNA
; ORGANISM: Paenibacillus polymyxa
US-09-784-554B-3

Alignment Scores:
Pred. No.: 1.298-10 Length: 4056
Score: 150.50 Matches: 27
Percent Similarity: 56.47% Conservative: 21
Best Local Similarity: 31.76% Mismatches: 36
Query Match: 32.02% Indels: 1
DB: 10 Gaps: 1

US-09-917-376-4 (1-89) x US-09-784-554B-3 (1-4056)
QY 4 GlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
Db 3595 GGATTGGTCTCCAGTATCGCACCGGATACAAATGTGACGACAACTCATTGAACCG 3654
QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
Db 3655 CATTTCCAAATTTAAATAAGGTACATCTCCGTACCGATCAACGAGTTGAAATTCGC 3714
QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63
Db 3715 TACTACTACACATCGACGGTGACCGTGACGAGCAG---ACATTCAACTGCGACTATGCGGTG 3771
QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
Db 3772 CTGAGCTGCTCGAAGCTGAATGGTAAAGCTGGTTAAATGGATAAAGCTGCAACCGGTGCT 3831
QY 84 AspThrTyrLeuGln 88
Db 3832 GATTATTATTGGAA 3846

RESULT 15
US-10-282-122A-15299
; Sequence 15299, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 15299
; LENGTH: 4452
; TYPE: DNA
; ORGANISM: Clostridium acetobutylicum
US-10-282-122A-15299

Alignment Scores:
Pred. No.: 1.04e-07 Length: 4452
Score: 130.50 Matches: 30
Percent Similarity: 55.79% Conservative: 23
Best Local Similarity: 31.58% Mismatches: 33
Query Match: 27.77% Indels: 9
DB: 13 Gaps: 2

US-09-917-376-4 (1-89) x US-10-282-122A-15299 (1-4452)
QY 2 SerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21
Db 100 AATTCGGGTGTTCAATACAAATTTGCTGATACAAATCTAGTACAAACCATGAACTATT 159
QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
Db 160 GCTCTTAATTTAAATATCAAAATAATATCTGGAGCACCTTTAGATTTAACAACCTTTAAA 219
QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61
Db 220 TTAAGATACTATTTTACAGCTGATGCTACTCAGGATGAAATTTTGG---TGCACCAT 276
QY 62 AlaAlaMetGlyCysGly-----AsnIleArgAlaSerPhe 73
Db 277 GCTGGTATGCTTAATGGTTATACTACCAACAATTTACAAGTAATGTAGTGGTACTTTT 336
QY 74 GlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 337 GTAGCTATGGTAATGCAACAGCTACTGCTGATCATTAICTTGAG 381

Search completed: May 12, 2004, 18:16:54
Job time : 817.654 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 12:10:32 ; Search time 54 Seconds  
(without alignments)  
465.680 Million cell updates/sec

Title: US-09-917-376-4  
Perfect score: 89  
Sequence: 1 VSGGVKQVKNNDSAPGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 1586107 seqs, 282547505 residues

Word size : 10

Total number of hits satisfying chosen parameters: 12

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database : A\_Geneseq\_29Jan04:\*  
1: geneseqp1980s:\*  
2: geneseqp1950s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	88	98.9	88	7	ADD22925
2	88	98.9	89	7	ADD22924
3	88	98.9	154	6	ABP73020
4	88	98.9	762	6	ABP73022
5	83	93.3	150	6	ABP73025
6	83	93.3	1228	6	ABP73029
7	71	79.8	153	6	ABP71658
8	71	79.8	1121	6	ABP71656
9	36	40.4	88	6	ABP73018
10	36	40.4	89	6	ABP73017
11	36	40.4	957	6	ABP73015
12	36	40.4	957	7	ADD22921

ALIGNMENTS

RESULT 1  
ADD22925  
ID ADD22925 standard; protein; 88 AA.  
XX  
AC ADD22925;  
XX  
DT 15-JAN-2004 (first entry)  
XX

DE Acidothermus cellulolyticus avicelase AvIII CBD III #2.  
XX enzyme; AvIII; cellulose reduction; agricultural biomass;  
KW municipal solid waste; glycoside hydrolase; avicelase.  
XX  
OS Acidothermus cellulolyticus.  
XX  
FN US2003108988-A1.  
XX  
PD 12-JUN-2003.  
XX  
PF 18-OCT-2002; 2002US-00155400.  
XX  
PR 28-JUL-2001; 2001US-00917376.  
XX  
PA (DING/) DING S.  
PA (ADNE/) ADNEY W S.  
PA (VINZ/) VINZANT T B.  
PA (HIMM/) HIMMEL M E.  
XX  
PI Ding S, Adney WS, Vinzant TB, Himmel ME;  
XX WPI; 2003-810853/76.  
DR  
XX  
PT New isolated thermal tolerant avicelase polynucleotide useful for  
PT detection of a polynucleotide encoding AvIII and for reducing cellulose  
PT in a starting material, e.g. municipal solid waste.  
XX  
PS Claim 16; SEQ ID NO 5; 29pp; English.  
XX  
CC The invention relates to an isolated polynucleotide molecule encoding a  
CC thermostable AvIII polypeptide. The polynucleotide is useful for  
CC detection of a polynucleotide encoding AvIII. The polynucleotide is  
CC useful for reducing cellulose in a starting material which involves  
CC administering to the starting material, e.g. agricultural biomass or  
CC municipal solid waste, a polypeptide molecule of the polynucleotide. The  
CC method further comprises administering a second polypeptide molecule  
CC chosen from the glycoside hydrolase family of proteins. The present  
CC sequence represents the amino acid sequence of Acidothermus  
CC cellulolyticus avicelase AvIII CBD III #2.  
XX  
SQ Sequence 88 AA;  
Query Match 98.9%; Score 88; DB 7; Length 88;  
Best Local Similarity 100.0%; Pred. No. 1.9e-81;  
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VSGGVKQVKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYVYFTRDGGSTLVNCD 60  
DB 1 VSGGVKQVKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYVYFTRDGGSTLVNCD 60  
QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
DB 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
RESULT 2  
ADD22924  
ID ADD22924 standard; protein; 89 AA.  
XX  
AC ADD22924;  
XX  
DT 15-JAN-2004 (first entry)  
XX  
DE Acidothermus cellulolyticus avicelase AvIII CBD III #1.  
KW enzyme; AvIII; cellulose reduction; agricultural biomass;  
KW municipal solid waste; glycoside hydrolase; avicelase.  
XX  
OS Acidothermus cellulolyticus.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 89

```
FT XX /label= Unknown
PN XX US2003108988-A1.
XX XX 12-JUN-2003.
XX XX 18-OCT-2002; 2002US-00155400.
XX XX 28-JUL-2001; 2001US-00917376.
XX XX (DING/) DING S.
PA (ADNE/) ADNEY W S.
PA (VINZ/) VINZANT T B.
PA (HIMM/) HIMMEL M E.
XX XX
PI Ding S, Adney WS, Vinzant TB, Himmel ME;
DR WPI; 2003-810853/76.
XX
PT New isolated thermal tolerant avicelase polynucleotide useful for
PT detection of a polynucleotide encoding Aviii and for reducing cellulose
PT in a starting material, e.g. municipal solid waste.
XX
PS Claim 16; SEQ ID NO 4; 29pp; English.
XX
CC The invention relates to an isolated polynucleotide molecule encoding a
CC thermostable Aviii polypeptide. The polynucleotide is useful for
CC detection of a polynucleotide encoding Aviii. The polynucleotide is
CC useful for reducing cellulose in a starting material which involves
CC administering to the starting material, e.g. agricultural biomass or
CC municipal solid waste, a polypeptide molecule of the polynucleotide. The
CC method further comprises administering a second polypeptide molecule
CC chosen from the glycoside hydrolase family of proteins. The present
CC sequence represents the amino acid sequence of Acidothermus
CC cellulolyticus avicelase Aviii CBD III #1.
XX
SQ Sequence 89 AA;
Query Match 98.9%; Score 88; DB 7; Length 89;
Best Local Similarity 100.0%; Pred. No. 1.9e-81;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VSGGVKQYKKNDSAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSSTLVYVNC 60
DB 1 VSGGVKQYKKNDSAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSSTLVYVNC 60
QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
DB 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
RESULT 3
ABP73020
ID ABP73020 standard; peptide; 154 AA.
XX
AC ABP73020;
XX
DT 03-JUN-2003 (first entry)
XX
DE Amino acid sequence of the ManA carbohydrate binding domain type III.
XX
KW ManA; mannanase A; glycoside hydrolase; enzyme; hemicellulose; sugar;
KW food; feed; paper pulp; biofuel; mannase.
XX
OS Acidothermus cellulolyticus.
XX
PN WO2003012110-A1.
XX
PD 13-FEB-2003.
XX
PF 28-JUL-2001; 2001WO-US023819.
XX
KW ManA; mannanase A; glycoside hydrolase; enzyme; hemicellulose; sugar;
KW food; feed; paper pulp; biofuel; mannase.
XX
OS Acidothermus cellulolyticus.
XX
PN WO2003012110-A1.
XX
PD 13-FEB-2003.
XX
PF 28-JUL-2001; 2001WO-US023819.
XX
PR 28-JUL-2001; 2001WO-US023819.
XX
PT Novel thermal tolerant mannanase A polypeptide derived from Acidothermus
PT cellulolyticus, useful for reducing hemicellulose in a starting material,
PT for processing of food, and as bulking agents in food stuffs.
XX
(MIDE ) MIDWEST RES INST.
Ding S, Adney WS, Vinzant TB, Himmel ME;
WPI; 2003-248182/24.
Novel thermal tolerant mannanase A polypeptide derived from Acidothermus
cellulolyticus, useful for reducing hemicellulose in a starting material,
for processing of food, and as bulking agents in food stuffs.
Claim 8; Page 7; 46pp; English.
The present sequence represents a fragment of ManA, a thermostable
mannanase A polypeptide derived from Acidothermus cellulolyticus. ManA is
a member of the glycoside hydrolase family of enzymes. ManA is useful for
reducing hemicellulose in a starting material to simpler carbohydrate
units, and ultimately to sugars which are useful in the food, feed, paper
pulp, and biofuels industries. It is useful for the processing of food
and in food stuffs as bulking agents, and for the degradation of mannanase.
ManA is also useful to raise polyclonal and monoclonal antibodies that
are useful in purifying ManA, or detecting ManA polypeptide expression,
and as well as reagent tools for characterizing the molecular actions of
CC ManA polypeptides
XX
SQ Sequence 154 AA;
Query Match 98.9%; Score 88; DB 6; Length 154;
Best Local Similarity 100.0%; Pred. No. 3.2e-81;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VSGGVKQYKKNDSAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSSTLVYVNC 60
DB 1 VSGGVKQYKKNDSAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSSTLVYVNC 60
QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
DB 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
RESULT 4
ABP73022
ID ABP73022 standard; protein; 762 AA.
XX
AC ABP73022;
XX
DT 03-JUN-2003 (first entry)
XX
DE Amino acid sequence of the ManA polypeptide.
XX
KW ManA; mannanase A; glycoside hydrolase; enzyme; hemicellulose; sugar;
KW food; feed; paper pulp; biofuel; mannase.
XX
OS Acidothermus cellulolyticus.
XX
PN WO2003012110-A1.
XX
PD 13-FEB-2003.
XX
PF 28-JUL-2001; 2001WO-US023819.
XX
PR 28-JUL-2001; 2001WO-US023819.
XX
PA (MIDE ) MIDWEST RES INST.
XX
PI Ding S, Adney WS, Vinzant TB, Himmel ME;
XX
DR WPI; 2003-248182/24.
DR N-PSDB; ABZ77633.
XX
PT Novel thermal tolerant mannanase A polypeptide derived from Acidothermus
PT cellulolyticus, useful for reducing hemicellulose in a starting material,
PT for processing of food, and as bulking agents in food stuffs.
```

XX PS Claim 3; Page 18-19; 46pp; English.

XX CC The present sequence represents Mana, a thermostable mannanase A

CC CC polypeptide derived from Acidothermus cellulolyticus. Mana is a member of

CC CC the glycoside hydrolase family of enzymes. Mana is useful for reducing

CC CC hemicellulose in a starting material to simpler carbohydrate units, and

CC CC ultimately to sugars which are useful in the food, feed, paper pulp, and

CC CC biofuels industries. It is useful for the processing of food and in food

CC CC stuffs as bulking agents, and for the degradation of mannan. Mana is

CC CC also useful to raise polyclonal and monoclonal antibodies that are useful

CC CC in purifying Mana, or detecting Mana polypeptide expression, and as well

CC CC as reagent tools for characterizing the molecular actions of Mana

XX CC polypeptides

XX SQ Sequence 762 AA;

Query Match 98.9%; Score 88; DB 6; Length 762;

Best Local Similarity 100.0%; Pred. No. 1.4e-80;

Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VSGGVKQYKNDSPAGDNOIKPGLQLVNTGSSVDLSVTVRYWFTDRDGGSSTLVYVNC 60

Db 455 VSGGVKQYKNDSPAGDNOIKPGLQLVNTGSSVDLSVTVRYWFTDRDGGSSTLVYVNC 514

Qy 61 WAAMGCGNIRASFGSVNPTPTADTYLQ 88

Db 515 WAAMGCGNIRASFGSVNPTPTADTYLQ 542

RESULT 5

ABP73025

ID ABP73025 standard; peptide, 150 AA.

XX AC ABP73025;

XX DT 03-JUN-2003 (first entry)

XX DE Amino acid sequence of the Guxa carbohydrate binding domain type III.

XX Guxa; cellulase; glycoside hydrolase; enzyme; cellulose; sugar; biofuel;

KW detergent; pulp processing; paper processing; feed processing; textile.

XX OS Acidothermus cellulolyticus.

XX WO2003012109-A1.

XX PD 13-FEB-2003.

XX PF 28-JUL-2001; 2001WO-US023817.

XX PR 28-JUL-2001; 2001WO-US023817.

XX PA (MIDE ) MIDWEST RES INST.

XX PI Ding S, Adney WS, Vinzant TB, Himmel ME, Decker SR;

XX WPI; 2003-239526/23.

XX Novel thermal tolerant Guxa polypeptide derived from Acidothermus

PT cellulolyticus, useful for reducing cellulose in a starting material, and

PT for the conversion of biomass to biofuels and biofuel additives.

XX Claim 6; Page 7; 47pp; English.

XX ABP73023-28 represent peptide fragments of a Guxa polypeptide. Guxa is a

CC thermostable cellulase, and is a member of the glycoside hydrolase family

CC of enzymes. Guxa is useful for reducing cellulose in a starting material

CC such as agricultural biomass to sugars. This is useful in biofuel

CC production. Guxa is also useful in the conversion of biomass to biofuels

CC and biofuel additives, in detergents, pulp and paper processing, food and

CC feed processing, and in textile process. Guxa is also useful for raising

CC polyclonal and monoclonal antibodies that are useful in purifying Guxa,

CC or detecting Guxa polypeptide expression, as well as reagent tools for

CC characterizing the molecular actions of Guxa polypeptides

XX SQ Sequence 150 AA;

Query Match 93.3%; Score 83; DB 6; Length 150;

Best Local Similarity 100.0%; Pred. No. 3.7e-76;

Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 KVOYKNDSPAGDNOIKPGLQLVNTGSSVDLSVTVRYWFTDRDGGSSTLVYVNC 65

Db 6 KVOYKNDSPAGDNOIKPGLQLVNTGSSVDLSVTVRYWFTDRDGGSSTLVYVNC 65

Qy 66 CGNIRASFGSVNPTPTADTYLQ 88

Db 66 CGNIRASFGSVNPTPTADTYLQ 88

RESULT 6

ABP73029

ID ABP73029 standard; protein, 1228 AA.

XX AC ABP73029;

XX DT 03-JUN-2003 (first entry)

XX DE Amino acid sequence of the Guxa polypeptide.

XX Guxa; cellulase; glycoside hydrolase; enzyme; cellulose; sugar; biofuel;

KW detergent; pulp processing; paper processing; feed processing; textile.

XX OS Acidothermus cellulolyticus.

XX WO2003012109-A1.

XX PD 13-FEB-2003.

XX PF 28-JUL-2001; 2001WO-US023817.

XX PR 28-JUL-2001; 2001WO-US023817.

XX PA (MIDE ) MIDWEST RES INST.

XX PI Ding S, Adney WS, Vinzant TB, Himmel ME, Decker SR;

XX WPI; 2003-239526/23.

XX N-PSDB; ABZ77634.

XX Novel thermal tolerant Guxa polypeptide derived from Acidothermus

PT cellulolyticus, useful for reducing cellulose in a starting material, and

PT for the conversion of biomass to biofuels and biofuel additives.

XX Claim 3; Page 19; 47pp; English.

XX The present sequence represents a Guxa polypeptide. Guxa is thermostable

CC cellulase, and is a member of the glycoside hydrolase family of enzymes.

CC Guxa is useful for reducing cellulose in a starting material such as

CC agricultural biomass to sugars. This is useful in biofuel production.

CC Guxa is also useful in the conversion of biomass to biofuels and biofuel

CC additives, in detergents, pulp and paper processing, food and feed

CC processing, and in textile process. Guxa is also useful for raising

CC polyclonal and monoclonal antibodies that are useful in purifying Guxa,

CC or detecting Guxa polypeptide expression, as well as reagent tools for

CC characterizing the molecular actions of Guxa polypeptides

XX SQ Sequence 1228 AA;

Query Match 93.3%; Score 83; DB 6; Length 1228;

Best Local Similarity 100.0%; Pred. No. 2.5e-75;

Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 KVOYKNDSPAGDNOIKPGLQLVNTGSSVDLSVTVRYWFTDRDGGSSTLVYVNC 65

Db 589 KVQYKNDSPGDQIKPGLQVNTGSSVDLSVTYRYWFTRDGSSSTLVNCDWAAMG 648  
 QY 66 CGNIRASFGSVNPATPTADTYLQ 88  
 |||||  
 Db 649 CGNIRASFGSVNPATPTADTYLQ 671

## RESULT 7

ABP71658  
 ID ABP71658 standard; protein; 153 AA.

XX AC ABP71658;

XX DT 29-MAY-2003 (first entry)

XX DE A. cellulolyticus Gux1 protein CBD\_III domain fragment.

XX KW Gux1; thermal tolerant; exoglucanase; glycoside hydrolase; cellulose;

XX KW biofuel; detergent; pulp; paper processing; feed processing; textile;

XX KW cellulase; enzyme.

XX OS Acidothermus cellulolyticus.

XX PN WO2003012095-A1.

XX PD 13-FEB-2003.

XX PF 28-JUL-2001; 2001WO-US023820.

XX PR 28-JUL-2001; 2001WO-US023820.

XX PA (MIDE ) MIDWEST RES INST.

XX PI Adney WS, Ding S, Vinzant TB, Himmel ME, Decker SR;

XX PI Lantz Mccarter S;

XX DR WPI; 2003-300494/29.

XX XX New thermal tolerant Gux1 peptide having specified amino acid sequence,

XX PT useful in the degradation of cellulose to biofuels.

XX PS Claim 2; Page 7; 44pp; English.

XX CC The invention relates to a thermal tolerant Gux1 peptide from A.

XX CC cellulolyticus. The Gux1 exoglucanase is a member of the glycoside

XX CC hydrolase family and comprises a catalytic domain GH48, carbohydrate

XX CC binding domain type III, and a carbohydrate binding domain type II. The

XX CC polypeptide is useful in the degradation of cellulose into biofuel, or

XX CC pulp and paper processing, food and feed processing, and in textile

XX CC processing. It can also be used alone or in combination with other

XX CC cellulase or glycoside hydrolases. The novel polypeptide generates

XX CC alternative cellulase enzymes capable of assisting in the commercial-

XX CC scale processing of cellulose to sugar for use in biofuel production. The

XX CC present sequence represents a A. cellulolyticus Gux1 cellulase CBD\_III

XX CC domain fragment

XX SQ Sequence 153 AA;

Query Match 79.8%; Score 71; DB 6; Length 153;

Best Local Similarity 100.0%; Pred. No. 6e-64;

Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 DNOIKPGLQVNTGSSVDLSVTYRYWFTRDGSSSTLVNCDWAAMCGNIRASFGSVN 77

Db 17 DNOIKPGLQVNTGSSVDLSVTYRYWFTRDGSSSTLVNCDWAAMCGNIRASFGSVN 76

QY 78 PATPTADTYLQ 88

Db 77 PATPTADTYLQ 87

## RESULT 8

QY 18 DNOIKPGLQVNTGSSVDLSVTYRYWFTRDGSSSTLVNCDWAAMCGNIRASFGSVN 77

Db 17 DNOIKPGLQVNTGSSVDLSVTYRYWFTRDGSSSTLVNCDWAAMCGNIRASFGSVN 76

QY 78 PATPTADTYLQ 88

Db 77 PATPTADTYLQ 87

## RESULT 9

QY 18 DNOIKPGLQVNTGSSVDLSVTYRYWFTRDGSSSTLVNCDWAAMCGNIRASFGSVN 77

Db 51 DNOIKPGLQVNTGSSVDLSVTYRYWFTRDGSSSTLVNCDWAAMCGNIRASFGSVN 110

QY 78 PATPTADTYLQ 88

Db 111 PATPTADTYLQ 121

ABP71656

ID ABP71656 standard; protein; 1121 AA.

XX AC ABP71656;

XX DT 29-MAY-2003 (first entry)

XX DE A. cellulolyticus Gux1 protein.

XX KW Gux1; thermal tolerant; exoglucanase; glycoside hydrolase; cellulose;

XX KW biofuel; detergent; pulp; paper processing; feed processing; textile;

XX KW cellulase; enzyme.

XX OS Acidothermus cellulolyticus.

XX FH Key Location/Qualifiers

FT Peptide 1..34

FT Protein /note= "potential signal peptide"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

FT /note= "mature protein"

XX	Acidothermus cellulolyticus.	OS
XX		XX
XX	Key	Location/Qualifiers
XX	Misc-difference	89
XX	/note= "unspecified residue"	
XX	WO2003012090-A2.	
XX		XX
XX	13-FEB-2003.	
XX		XX
XX	28-JUL-2001; 2001WO-US023818.	
XX		XX
XX	28-JUL-2001; 2001WO-US023818.	
XX	(MIDE ) MIDWEST RES INST.	
XX		XX
XX	Ding S, Adney WS, Vinzant TB, Himmel ME;	
XX	WPI; 2003-248177/24.	
XX		XX
XX	New thermostable AvIII peptide from Acidothermus cellulolyticus, useful	
XX	for degradation of cellulose or in generating anti-AvIII antibodies for	
XX	purifying recombinant AvIII polypeptides from genetically engineered	
XX	host cells.	
XX		XX
XX	Claim 6; Page 8; 4app; English.	
XX		XX
XX	The present sequence is derived from a thermostable avicelase, designated	
XX	AvIII. AvIII is a member of the glycoside hydrolase family of enzymes,	
XX	and is a cellulase. AvIII is useful in the conversion of biomass to	
XX	biofuels and biofuel additives. It may be useful in the production of	
XX	detergents, pulp and paper processing, food and feed processing and in	
XX	textile processes. The thermostable AvIII peptide is useful in the	
XX	degradation of cellulose, and in generating specific anti-AvIII	
XX	antibodies that are useful in purifying recombinant AvIII polypeptides	
XX	from genetically engineered host cells, in detecting AvIII polypeptide	
XX	expression, as well as a reagent tool for characterizing the molecular	
XX	actions of the polypeptide. The AvIII polynucleotide is useful as a	
XX	source of probes or primers in various diagnostic assays	
XX		XX
XX	Sequence 89 AA;	
XX		XX
XX	Query Match	40.4%; Score 36; DB 6; Length 89;
XX	Best Local Similarity	100.0%; Pred. No. 1.3e-28;
XX	Matches	36; Conservative 0; Mismatches 0; Indels 0; Gaps 0
XX		
QY	28 VNTGSSVDLSTVTYRYWFTTRDGGSTLVYNCWAA 63	
DB	28 VNTGSSVDLSTVTYRYWFTTRDGGSTLVYNCWAA 63	
XX		XX
XX	RESULT 11	
XX	ABP73015	
XX	ID ABP73015 standard; protein; 957 AA.	
XX		XX
XX	ABP73015;	
XX		XX
XX	03-JUN-2003 (first entry)	
XX		XX
XX	Amino acid sequence of the avicelase AvIII.	
XX		XX
XX	Avicelase; AvIII; glycoside hydrolase; enzyme; cellulase; biofuel;	
XX	detergent; pulp processing; paper processing; feed processing; textile;	
XX	cellulose.	
XX		XX
XX	Acidothermus cellulolyticus.	
XX		XX
XX	Key	Location/Qualifiers
XX	Misc-difference	957
XX	/note= "unspecified residue encoded by N"	
XX	WO2003012090-A2.	

XX 13-FEB-2003.  
 PD 28-JUL-2001; 2001WO-US023818.  
 XX 28-JUL-2001; 2001WO-US023818.  
 PR (MIDE ) MIDWEST RES INST.  
 XX Ding S, Adney WS, Vinzant TB, Himmel ME;  
 PI WPI; 2003-248177/24.  
 XX N-PSDB; ABZ77632.  
 DR New thermostable AvIII peptide from Acidothermus cellulolyticus, useful  
 XX for degradation of cellulose or in generating anti-AvIII antibodies for  
 PT purifying recombinant AvIII polypeptides from genetically engineered  
 PT host cells.  
 XX Claim 2; Page 20; 44pp; English.  
 PS The present sequence represents a thermostable avicelase polypeptide,  
 CC designated AvIII. AvIII is a member of the glycoside hydrolase family  
 CC of enzymes, and is a cellulase. AvIII is useful in the conversion of  
 CC biomass to biofuels and biofuel additives. It may be useful in the  
 CC production of detergents, pulp and paper processing, food and feed  
 CC processing and in textile processes. The thermostable AvIII peptide is  
 CC useful in the degradation of cellulose, and in generating specific anti-  
 CC AvIII antibodies that are useful in purifying recombinant AvIII  
 CC polypeptides from genetically engineered host cells, in detecting AvIII  
 CC polypeptide expression, as well as a reagent tool for characterizing the  
 CC molecular actions of the polypeptide. The AvIII polynucleotide is useful  
 CC as a source of probes or primers in various diagnostic assays  
 XX Sequence 957 AA;  
 SQ

Query Match 40.4%; Score 36; DB 6; Length 957;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-27;  
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 28 VNTGSSVDLSVTVTYVWFTRDGSGSTLVYNCDAWAA 63  
 DB 896 VNTGSSVDLSVTVTYVWFTRDGSGSTLVYNCDAWAA 931

RESULT 12  
 ADD22921  
 ID ADD22921 standard; protein; 957 AA.  
 XX  
 AC ADD22921;  
 XX  
 DT 15-JAN-2004 (first entry)  
 XX  
 DE Acidothermus cellulolyticus avicelase AvIII.  
 XX  
 KW enzyme; AvIII; cellulose reduction; agricultural biomass;  
 KW municipal solid waste; glycoside hydrolase; avicelase.  
 XX  
 OS Acidothermus cellulolyticus.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 957  
 FT /label= Unknown  
 FT /note= "Encoded by N"  
 XX  
 PN US2003108988-A1.  
 XX  
 PD 12-JUN-2003.  
 XX  
 PF 18-OCT-2002; 2002US-00155400.  
 XX  
 PR 28-JUL-2001; 2001US-00917376.  
 XX

PA (DING/) DING S.  
 PA (ADNE/) ADNEY W S.  
 PA (VINZ/) VINZANT T B.  
 PA (HIMM/) HIMMEL M E.  
 XX  
 PI Ding S, Adney WS, Vinzant TB, Himmel ME;  
 XX WPI; 2003-810853/76.  
 DR N-PSDB; ADD22922.  
 XX  
 PT New isolated thermal tolerant avicelase polynucleotide useful for  
 PT detection of a polynucleotide encoding AvIII and for reducing cellulose  
 PT in a starting material, e.g. municipal solid waste.  
 XX  
 PS Claim 16; SEQ ID NO 1; 29pp; English.  
 XX  
 CC The invention relates to an isolated polynucleotide molecule encoding a  
 CC thermostable AvIII polypeptide. The polynucleotide is useful for  
 CC detection of a polynucleotide encoding AvIII. The polynucleotide is  
 CC useful for reducing cellulose in a starting material which involves  
 CC administering to the starting material, e.g. agricultural biomass or  
 CC municipal solid waste, a polypeptide molecule of the polynucleotide. The  
 CC method further comprises administering a second polypeptide molecule  
 CC chosen from the glycoside hydrolase family of proteins. The present  
 CC sequence represents the amino acid sequence of Acidothermus  
 CC cellulolyticus avicelase AvIII.  
 XX Sequence 957 AA;  
 SQ

Query Match 40.4%; Score 36; DB 7; Length 957;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-27;  
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 28 VNTGSSVDLSVTVTYVWFTRDGSGSTLVYNCDAWAA 63  
 DB 896 VNTGSSVDLSVTVTYVWFTRDGSGSTLVYNCDAWAA 931

Search completed: May 11, 2004, 12:26:52  
 Job time : 55 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 12:23:57 ; Search time 21 Seconds  
(without alignments)  
407.669 Million cell updates/sec

Title: US-09-917-376-4  
Perfect score: 89  
Sequence: 1 VSGGVKQYKXNDSAPGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 283366 seqs, 96191526 residues

Word size : 10

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : PIR 78.\*

- 1: PIR1.\*
- 2: PIR2.\*
- 3: PIR3.\*
- 4: PIR4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
-----					

No matches found

Search completed: May 11, 2004, 12:28:43  
Job time : 21 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 12:11:27 ; Search time 15 Seconds  
(without alignments)  
308.950 Million cell updates/sec

Title: US-09-917-376-4  
Perfect score: 89  
Sequence: 1 VSGGVKQYKNKNSAPGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 141681 seqs, 52070155 residues

Word size : 10

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
-----					

No matches found

Search completed: May 11, 2004, 12:27:19  
Job time : 16 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 12:23:22 ; Search time 39 Seconds  
(without alignments)  
720.029 Million cell updates/sec

Title: US-09-917-376-4  
Perfect score: 89  
Sequence: 1 VSGGVKYKXNDSPAGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 1017041 seqs, 315518202 residues

Word size : 10

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

- SPTREMBL 25:\*
- 1: sp\_archaea:\*
  - 2: sp\_bacteria:\*
  - 3: sp\_fungi:\*
  - 4: sp\_human:\*
  - 5: sp\_invertebrate:\*
  - 6: sp\_mammal:\*
  - 7: sp\_mhc:\*
  - 8: sp\_organelle:\*
  - 9: sp\_phage:\*
  - 10: sp\_plant:\*
  - 11: sp\_rodent:\*
  - 12: sp\_virus:\*
  - 13: sp\_vertebrate:\*
  - 14: sp\_unclassified:\*
  - 15: sp\_virus:\*
  - 16: sp\_bacteriap:\*
  - 17: sp\_archheap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Score	Match	Length	DB	ID	Description
--------	-------	-------	--------	----	----	-------------

No matches found

Search completed: May 11, 2004, 12:28:10  
Job time : 39 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 12:25:52 ; Search time 23 Seconds  
(without alignments)  
199.770 Million cell updates/sec

Title: US-09-917-376-4  
Perfect score: 89  
Sequence: 1 VSGGVKVKYKNDSPAGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 389414 seqs, 51625971 residues

Word size : 10

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
---------------	-------	----------------	--------	----	-------------

No matches found

Search completed: May 11, 2004, 12:29:19  
Job time : 23 secs

```

Query Match      98.9%; Score 88; DB 14; Length 88;
Best Local Similarity 100.0%; Pred. No. 1.7e-78;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VSGGVKYYKNNDSPAGDNOIKPGLQLVNTGSSSDLTSTVRYWFRDGGSSSTLVNCD 60
Db 1 VSGGVKYYKNNDSPAGDNOIKPGLQLVNTGSSSDLTSTVRYWFRDGGSSSTLVNCD 60

```

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
Db 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88

## RESULT 3

US-09-917-376-4  
; Sequence 4, Application US/09917376  
; Publication No. US20040038334A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS  
; FILE REFERENCE: 40197.4US01  
; CURRENT APPLICATION NUMBER: US/09/917,376  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 89  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
; FEATURE:  
; NAME/KEY: MOD\_RES  
; LOCATION: (89)  
; OTHER INFORMATION: Any amino acid  
; OTHER INFORMATION: Carbohydrate binding domain  
US-09-917-376-4

Query Match 98.9%; Score 88; DB 12; Length 89;  
Best Local Similarity 100.0%; Pred. No. 1.7e-78;  
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKVOYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLVNCD 60  
Db 1 VSGGVKVOYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLVNCD 60

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
Db 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88

## RESULT 4

US-10-155-400-4  
; Sequence 4, Application US/10155400  
; Publication No. US20030108988A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS  
; FILE REFERENCE: NREL 01-36A  
; CURRENT APPLICATION NUMBER: US/10/155,400  
; CURRENT FILING DATE: 2002-10-22  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 89  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
; FEATURE:  
; NAME/KEY: MOD\_RES  
; LOCATION: (89)  
; OTHER INFORMATION: Any amino acid  
; FEATURE:  
; OTHER INFORMATION: Carbohydrate binding domain  
US-10-155-400-4

Query Match 98.9%; Score 88; DB 14; Length 89;  
Best Local Similarity 100.0%; Pred. No. 1.7e-78;  
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKVOYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLVNCD 60  
Db 1 VSGGVKVOYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLVNCD 60

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
Db 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88

## RESULT 5

US-09-917-378-4  
; Sequence 4, Application US/09917378  
; Publication No. US20030119093A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: DECKER, STEPHEN R.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT MANNANASE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOXYLICUS  
; FILE REFERENCE: 40197.7US01  
; CURRENT APPLICATION NUMBER: US/09/917,378  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 154  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Carbohydrate  
; OTHER INFORMATION: binding domain  
US-09-917-378-4

Query Match 98.9%; Score 88; DB 10; Length 154;  
Best Local Similarity 100.0%; Pred. No. 2.7e-78;  
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKVOYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLVNCD 60  
Db 1 VSGGVKVOYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLVNCD 60

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
Db 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88

## RESULT 6

US-09-917-378-1  
; Sequence 1, Application US/09917378  
; Publication No. US20030119093A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: DECKER, STEPHEN R.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT MANNANASE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOXYLICUS  
; FILE REFERENCE: 40197.7US01  
; CURRENT APPLICATION NUMBER: US/09/917,378  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 762  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus

```
US-09-917-378-1
Query Match      98.9%; Score 88; DB 10; Length 762;
Best Local Similarity 100.0%; Pred. No. 1.1e-77;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VSGGVKQYKNDSPAGDNOIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSSTLVNCD 60
Db 455 VSGGVKQYKNDSPAGDNOIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSSTLVNCD 514

Qy 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
Db 515 WAAMGCGNIRASFGSVNPATPTADTYLQ 542

RESULT 7
US-09-917-384-5
; Sequence 5, Application US/09917384
; Publication No. US20030096342A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,384
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 150
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
US-09-917-384-5

Query Match      93.3%; Score 83; DB 10; Length 150;
Best Local Similarity 100.0%; Pred. No. 2.1e-73;
Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 KQYKNDSPAGDNOIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSSTLVNCDWAAMG 65
Db 6 KQYKNDSPAGDNOIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSSTLVNCDWAAMG 65

Qy 66 CGNIRASFGSVNPATPTADTYLQ 88
Db 66 CGNIRASFGSVNPATPTADTYLQ 88

RESULT 8
US-09-917-383-5
; Sequence 5, Application US/09917383
; Publication No. US20030104522A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,383
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 150
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
US-09-917-383-5

Query Match      93.3%; Score 83; DB 10; Length 1043;
Best Local Similarity 100.0%; Pred. No. 1.1e-72;
Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 KQYKNDSPAGDNOIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSSTLVNCDWAAMG 65
Db 482 KQYKNDSPAGDNOIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSSTLVNCDWAAMG 541

Qy 66 CGNIRASFGSVNPATPTADTYLQ 88
Db 542 CGNIRASFGSVNPATPTADTYLQ 564

RESULT 9
US-09-917-384-6
; Sequence 6, Application US/09917384
; Publication No. US20030096342A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,384
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1043
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
US-09-917-384-6

Query Match      93.3%; Score 83; DB 10; Length 1043;
Best Local Similarity 100.0%; Pred. No. 1.1e-72;
Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 KQYKNDSPAGDNOIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSSTLVNCDWAAMG 65
Db 482 KQYKNDSPAGDNOIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSSTLVNCDWAAMG 541

Qy 66 CGNIRASFGSVNPATPTADTYLQ 88
Db 542 CGNIRASFGSVNPATPTADTYLQ 564

RESULT 10
US-09-917-383-6
; Sequence 6, Application US/09917383
; Publication No. US20030104522A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,383
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1043
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
US-09-917-383-6
```

; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 1043  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Segment of  
; OTHER INFORMATION: Guxa  
US-09-917-383-6

Query Match 93.3%; Score 83; DB 10; Length 1043;  
Best Local Similarity 100.0%; Pred. No. 1.1e-72;  
Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 6 KVOYKNDSPAGDNOIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSSTLVYNCDAWAMG 65  
DB 482 KVOYKNDSPAGDNOIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSSTLVYNCDAWAMG 541  
QY 66 CGNIRASFGSVNPATPTADTYLQ 88  
DB 542 CGNIRASFGSVNPATPTADTYLQ 564

RESULT 11  
US-09-917-384-1  
; Sequence 1, Application US/09917384  
; Publication No. US20030096342A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: DECKER, STEPHEN R.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOLYTICUS  
; FILE REFERENCE: 40170.6US01  
; CURRENT APPLICATION NUMBER: US/09/917,384  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 1228  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Segment of  
; OTHER INFORMATION: Guxa  
US-09-917-384-1

Query Match 93.3%; Score 83; DB 10; Length 1228;  
Best Local Similarity 100.0%; Pred. No. 1.3e-72;  
Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 6 KVOYKNDSPAGDNOIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSSTLVYNCDAWAMG 65  
DB 599 KVOYKNDSPAGDNOIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSSTLVYNCDAWAMG 648  
QY 66 CGNIRASFGSVNPATPTADTYLQ 88  
DB 649 CGNIRASFGSVNPATPTADTYLQ 671

RESULT 12  
US-09-917-383-1  
; Sequence 1, Application US/09917383  
; Publication No. US20030104522A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: DECKER, STEPHEN R.

; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOLYTICUS  
; FILE REFERENCE: 40170.6US01  
; CURRENT APPLICATION NUMBER: US/09/917,383  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 1228  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Segment of  
; OTHER INFORMATION: Guxa  
US-09-917-383-1

Query Match 93.3%; Score 83; DB 10; Length 1228;  
Best Local Similarity 100.0%; Pred. No. 1.3e-72;  
Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 6 KVOYKNDSPAGDNOIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSSTLVYNCDAWAMG 65  
DB 589 KVOYKNDSPAGDNOIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSSTLVYNCDAWAMG 648  
QY 66 CGNIRASFGSVNPATPTADTYLQ 88  
DB 649 CGNIRASFGSVNPATPTADTYLQ 671

RESULT 13  
US-09-917-376-1  
; Sequence 1, Application US/09917376  
; Publication No. US20040038334A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOLYTICUS  
; FILE REFERENCE: 40197.4US01  
; CURRENT APPLICATION NUMBER: US/09/917,376  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 957  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
; FEATURE:  
; NAME/KEY: MOD RES  
; LOCATION: (957)  
; OTHER INFORMATION: Any amino acid  
US-09-917-376-1

Query Match 40.4%; Score 36; DB 12; Length 957;  
Best Local Similarity 100.0%; Pred. No. 1.3e-26;  
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 28 VNTGSSVDLSTVTYRYWFTRDGSSSTLVYNCDAWAA 63  
DB 896 VNTGSSVDLSTVTYRYWFTRDGSSSTLVYNCDAWAA 931

RESULT 14  
US-10-155-400-1  
; Sequence 1, Application US/10155400  
; Publication No. US20030108988A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.



Qy 28 VNTGSSVDLSTVTVRYMFTRDGGSSTLVYNCDAW 63  
|||  
Db 896 VNTGSSVDLSTVTVRYMFTRDGGSSTLVYNCDAW 931

Search completed: May 11, 2004, 12:33:59  
Job time : 42 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 12:26:57 ; Search time 182 Seconds  
(without alignments)  
477.301 Million cell updates/sec

Title: US-09-917-376-4

Perfect score: 89

Sequence: 1 VSGGVKQYKNNDSAPGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 6019581 seqs, 976053577 residues

Word size : 10

Total number of hits satisfying chosen parameters: 16

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Pending Patents\_AA\_Main.\*

- 1: /cgn2\_6/ptodata/2/paa/PCTUS\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/2/paa/US06\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/2/paa/US07\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/2/paa/US080\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/2/paa/US081\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/2/paa/US082\_COMB.pep.\*
- 7: /cgn2\_6/ptodata/2/paa/US083\_COMB.pep.\*
- 8: /cgn2\_6/ptodata/2/paa/US084\_COMB.pep.\*
- 9: /cgn2\_6/ptodata/2/paa/US085\_COMB.pep.\*
- 10: /cgn2\_6/ptodata/2/paa/US086\_COMB.pep.\*
- 11: /cgn2\_6/ptodata/2/paa/US087\_COMB.pep.\*
- 12: /cgn2\_6/ptodata/2/paa/US088\_COMB.pep.\*
- 13: /cgn2\_6/ptodata/2/paa/US089\_COMB.pep.\*
- 14: /cgn2\_6/ptodata/2/paa/US090\_COMB.pep.\*
- 15: /cgn2\_6/ptodata/2/paa/US091\_COMB.pep.\*
- 16: /cgn2\_6/ptodata/2/paa/US092\_COMB.pep.\*
- 17: /cgn2\_6/ptodata/2/paa/US093\_COMB.pep.\*
- 18: /cgn2\_6/ptodata/2/paa/US094\_COMB.pep.\*
- 19: /cgn2\_6/ptodata/2/paa/US095\_COMB.pep.\*
- 20: /cgn2\_6/ptodata/2/paa/US096\_COMB.pep.\*
- 21: /cgn2\_6/ptodata/2/paa/US097A\_COMB.pep.\*
- 22: /cgn2\_6/ptodata/2/paa/US097B\_COMB.pep.\*
- 23: /cgn2\_6/ptodata/2/paa/US098\_COMB.pep.\*
- 24: /cgn2\_6/ptodata/2/paa/US099A\_COMB.pep.\*
- 25: /cgn2\_6/ptodata/2/paa/US099B\_COMB.pep.\*
- 26: /cgn2\_6/ptodata/2/paa/US100\_COMB.pep.\*
- 27: /cgn2\_6/ptodata/2/paa/US101\_COMB.pep.\*
- 28: /cgn2\_6/ptodata/2/paa/US102\_COMB.pep.\*
- 29: /cgn2\_6/ptodata/2/paa/US103\_COMB.pep.\*
- 30: /cgn2\_6/ptodata/2/paa/US104\_COMB.pep.\*
- 31: /cgn2\_6/ptodata/2/paa/US106\_COMB.pep.\*
- 32: /cgn2\_6/ptodata/2/paa/US107\_COMB.pep.\*
- 33: /cgn2\_6/ptodata/2/paa/US60\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match Length	DB ID	Description
------------	-------	--------------	-------	-------------

RESULT 1

US-09-917-376-5

; Sequence 5, Application US/09917376

; GENERAL INFORMATION:

; APPLICANT: DING, SHI-YOU

; APPLICANT: ADNEY, WILLIAM S.

; APPLICANT: VINZANT, TODD B.

; APPLICANT: HIMMEL, MICHAEL E.

; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS

; FILE OF INVENTION: CELLULOYTICUS

; FILE REFERENCE: 40197.4US01

; CURRENT APPLICATION NUMBER: US/09/917,376

; CURRENT FILING DATE: 2001-07-28

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: Patent in Ver. 2.1

; SEQ ID NO 5

; LENGTH: 88

; TYPE: PRT

; ORGANISM: Acidothermus cellulolyticus

; FEATURE:

; OTHER INFORMATION: Carbohydrate binding domain

US-09-917-376-5

Query Match Similarity 98.9%; Score 88; DB 24; Length 88;

Best Local Similarity 100.0%; Pred. No. 8e-81; Mismatches 0; Indels 0; Gaps 0;

Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VSGGVKQYKNNDSAPGDNQIKPGLQVNTGSSVDLSITVTVYVWTRDGGSTLYNCD 60

Db 1 VSGGVKQYKNNDSAPGDNQIKPGLQVNTGSSVDLSITVTVYVWTRDGGSTLYNCD 60

Qy 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88

Db 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88

RESULT 2

US-10-155-400-5

; Sequence 5, Application US/10155400

; GENERAL INFORMATION:

; APPLICANT: DING, SHI-YOU

; APPLICANT: ADNEY, WILLIAM S.

; APPLICANT: VINZANT, TODD B.

; APPLICANT: HIMMEL, MICHAEL E.

; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS

; FILE OF INVENTION: CELLULOYTICUS

; FILE REFERENCE: NREL 01-36A

; CURRENT APPLICATION NUMBER: US/10/155,400

; CURRENT FILING DATE: 2002-10-22

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: Patent in Ver. 2.1

; SEQ ID NO 5

```
; LENGTH: 88
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; OTHER INFORMATION: Carbohydrate binding domain
US-10-155-400-5

Query Match      98.9%; Score 88; DB 27; Length 88;
Best Local Similarity 100.0%; Pred. No. 8e-81;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVQKNDSPAGDNQIKPGLQLVNTGSSSVLDLSTVTVRYWFTTRDGGSSTLVYVNC 60
   |||||
Db 1 VSGGVKQVQKNDSPAGDNQIKPGLQLVNTGSSSVLDLSTVTVRYWFTTRDGGSSTLVYVNC 60
   |||||

QY 61 WAAMCGGNIRASFGSVNPATPTADTYLQ 88
   |||||
Db 61 WAAMCGGNIRASFGSVNPATPTADTYLQ 88
   |||||

RESULT 3
US-09-917-376-4
; Sequence 4, Application US/09917376
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOYTICUS
; FILE REFERENCE: 40197.4US01
; CURRENT APPLICATION NUMBER: US/09/917,376
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 89
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; NAME/KEY: MOD.RES
; LOCATION: (89)
; OTHER INFORMATION: Any amino acid
; OTHER INFORMATION: Carbohydrate binding domain
US-09-917-376-4

Query Match      98.9%; Score 88; DB 24; Length 89;
Best Local Similarity 100.0%; Pred. No. 8.1e-81;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVQKNDSPAGDNQIKPGLQLVNTGSSSVLDLSTVTVRYWFTTRDGGSSTLVYVNC 60
   |||||
Db 1 VSGGVKQVQKNDSPAGDNQIKPGLQLVNTGSSSVLDLSTVTVRYWFTTRDGGSSTLVYVNC 60
   |||||

QY 61 WAAMCGGNIRASFGSVNPATPTADTYLQ 88
   |||||
Db 61 WAAMCGGNIRASFGSVNPATPTADTYLQ 88
   |||||

RESULT 4
US-10-155-400-4
; Sequence 4, Application US/10155400
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOYTICUS
; FILE REFERENCE: NREL 01-36A
; CURRENT APPLICATION NUMBER: US/10/155,400
; CURRENT FILING DATE: 2002-10-22
; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 89
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; NAME/KEY: MOD.RES
; LOCATION: (89)
; OTHER INFORMATION: Any amino acid
; OTHER INFORMATION: Carbohydrate binding domain
US-09-917-376-4

Query Match      98.9%; Score 88; DB 24; Length 89;
Best Local Similarity 100.0%; Pred. No. 8.1e-81;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVQKNDSPAGDNQIKPGLQLVNTGSSSVLDLSTVTVRYWFTTRDGGSSTLVYVNC 60
   |||||
Db 1 VSGGVKQVQKNDSPAGDNQIKPGLQLVNTGSSSVLDLSTVTVRYWFTTRDGGSSTLVYVNC 60
   |||||

QY 61 WAAMCGGNIRASFGSVNPATPTADTYLQ 88
   |||||
Db 61 WAAMCGGNIRASFGSVNPATPTADTYLQ 88
   |||||

RESULT 6
US-09-917-378-1
; Sequence 1, Application US/09917378
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT MANNANASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOYTICUS
; FILE REFERENCE: 40197.7US01
; CURRENT APPLICATION NUMBER: US/09/917,378
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 154
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Carbohydrate
; OTHER INFORMATION: binding domain
US-09-917-378-4

Query Match      98.9%; Score 88; DB 24; Length 154;
Best Local Similarity 100.0%; Pred. No. 1.3e-80;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVQKNDSPAGDNQIKPGLQLVNTGSSSVLDLSTVTVRYWFTTRDGGSSTLVYVNC 60
   |||||
Db 1 VSGGVKQVQKNDSPAGDNQIKPGLQLVNTGSSSVLDLSTVTVRYWFTTRDGGSSTLVYVNC 60
   |||||

QY 61 WAAMCGGNIRASFGSVNPATPTADTYLQ 88
   |||||
Db 61 WAAMCGGNIRASFGSVNPATPTADTYLQ 88
   |||||
```

; TITLE OF INVENTION: THERMAL TOLERANT MANNANASE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOLYTICUS  
; FILE REFERENCE: 40197.7US01  
; CURRENT APPLICATION NUMBER: US/09/917,378  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 762  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
US-09-917-378-1

Query Match 98.9%; Score 88; DB 24; Length 762;  
Best Local Similarity 100.0%; Pred. No. 5.8e-80;  
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 VSGGVKQYKNDSPAGDQNIKPGQLQVNTGSSVDLSVTVRYWFTRDGSSSTLVNCD 60  
Db 455 VSGGVKQYKNDSPAGDQNIKPGQLQVNTGSSVDLSVTVRYWFTRDGSSSTLVNCD 514  
Qy 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88  
Db 515 WAAMGCGNIRASFGSVNPATPTADTYLQ 542

RESULT 7  
US-09-917-383-5  
; Sequence 5, Application US/09917383  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: DECKER, STEPHEN R.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOLYTICUS  
; FILE REFERENCE: 40170.6US01  
; CURRENT APPLICATION NUMBER: US/09/917,383  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 150  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Segment of  
; OTHER INFORMATION: Guxa  
US-09-917-383-5

Query Match 93.3%; Score 83; DB 24; Length 150;  
Best Local Similarity 100.0%; Pred. No. 1.6e-75;  
Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 6 KVQYKNDSPAGDQNIKPGQLQVNTGSSVDLSVTVRYWFTRDGSSSTLVNCDWAAMG 65  
Db 6 KVQYKNDSPAGDQNIKPGQLQVNTGSSVDLSVTVRYWFTRDGSSSTLVNCDWAAMG 65  
Qy 66 CGNIRASFGSVNPATPTADTYLQ 88  
Db 66 CGNIRASFGSVNPATPTADTYLQ 88

RESULT 8  
US-09-917-384-5  
; Sequence 5, Application US/09917384  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: DECKER, STEPHEN R.  
; APPLICANT: HIMMEL, MICHAEL E.

; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOLYTICUS  
; FILE REFERENCE: 40170.6US01  
; CURRENT APPLICATION NUMBER: US/09/917,384  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 150  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Segment of  
; OTHER INFORMATION: Guxa  
US-09-917-384-5

Query Match 93.3%; Score 83; DB 24; Length 150;  
Best Local Similarity 100.0%; Pred. No. 1.6e-75;  
Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 6 KVQYKNDSPAGDQNIKPGQLQVNTGSSVDLSVTVRYWFTRDGSSSTLVNCDWAAMG 65  
Db 6 KVQYKNDSPAGDQNIKPGQLQVNTGSSVDLSVTVRYWFTRDGSSSTLVNCDWAAMG 65  
Qy 66 CGNIRASFGSVNPATPTADTYLQ 88  
Db 66 CGNIRASFGSVNPATPTADTYLQ 88

RESULT 9  
US-09-917-383-6  
; Sequence 6, Application US/09917383  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: DECKER, STEPHEN R.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOLYTICUS  
; FILE REFERENCE: 40170.6US01  
; CURRENT APPLICATION NUMBER: US/09/917,383  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 1043  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Segment of  
; OTHER INFORMATION: Guxa  
US-09-917-383-6

Query Match 93.3%; Score 83; DB 24; Length 1043;  
Best Local Similarity 100.0%; Pred. No. 9.4e-75;  
Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 6 KVQYKNDSPAGDQNIKPGQLQVNTGSSVDLSVTVRYWFTRDGSSSTLVNCDWAAMG 65  
Db 482 KVQYKNDSPAGDQNIKPGQLQVNTGSSVDLSVTVRYWFTRDGSSSTLVNCDWAAMG 541  
Qy 66 CGNIRASFGSVNPATPTADTYLQ 88  
Db 542 CGNIRASFGSVNPATPTADTYLQ 564

RESULT 10  
US-09-917-384-6  
; Sequence 6, Application US/09917384  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.

```
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,384
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1043
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
US-09-917-384-6

Query Match          93.3%; Score 83; DB 24; Length 1043;
Best Local Similarity 100.0%; Pred. No. 9.4e-75;
Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 KVQYKNDSAPGDNQIKPGLQLVNTGSSVDLSVTYVYWFTRDGGSSTLVYNCDDWAAMG 65
Db 482 KVQYKNDSAPGDNQIKPGLQLVNTGSSVDLSVTYVYWFTRDGGSSTLVYNCDDWAAMG 541

Qy 66 CGNIRASFGSVNPATPTADTYLQ 88
Db 542 CGNIRASFGSVNPATPTADTYLQ 564

RESULT 11
US-09-917-383-1
; Sequence 1, Application US/09917383
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,383
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1228
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
US-09-917-383-1

Query Match          93.3%; Score 83; DB 24; Length 1228;
Best Local Similarity 100.0%; Pred. No. 1.1e-74;
Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 KVQYKNDSAPGDNQIKPGLQLVNTGSSVDLSVTYVYWFTRDGGSSTLVYNCDDWAAMG 65
Db 589 KVQYKNDSAPGDNQIKPGLQLVNTGSSVDLSVTYVYWFTRDGGSSTLVYNCDDWAAMG 648

Qy 66 CGNIRASFGSVNPATPTADTYLQ 88
Db 649 CGNIRASFGSVNPATPTADTYLQ 671

RESULT 12
US-09-917-384-1
; Sequence 1, Application US/09917384
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,384
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 4
; LENGTH: 153
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
US-09-917-384-1

Query Match          79.8%; Score 71; DB 24; Length 153;
Best Local Similarity 100.0%; Pred. No. 2.5e-63;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 DNOIKPGLQLVNTGSSVDLSVTYVYWFTRDGGSSTLVYNCDDWAAMGCGNIRASFGSVN 77
Db 17 DNOIKPGLQLVNTGSSVDLSVTYVYWFTRDGGSSTLVYNCDDWAAMGCGNIRASFGSVN 76

Qy 78 PATPTADTYLQ 88
Db 77 PATPTADTYLQ 87
```

```
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,384
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1228
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
US-09-917-384-1

Query Match          93.3%; Score 83; DB 24; Length 1228;
Best Local Similarity 100.0%; Pred. No. 1.1e-74;
Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 KVQYKNDSAPGDNQIKPGLQLVNTGSSVDLSVTYVYWFTRDGGSSTLVYNCDDWAAMG 65
Db 589 KVQYKNDSAPGDNQIKPGLQLVNTGSSVDLSVTYVYWFTRDGGSSTLVYNCDDWAAMG 648

Qy 66 CGNIRASFGSVNPATPTADTYLQ 88
Db 649 CGNIRASFGSVNPATPTADTYLQ 671

RESULT 13
US-09-917-384A-4
; Sequence 4, Application US/09917384A
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: MCCARTER, SUZANNE
; APPLICANT: HIMMEL, MICHAEL E.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: VINZANT, TODD B.
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: NREL 01-38
; CURRENT APPLICATION NUMBER: US/09/917,384A
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 4
; LENGTH: 153
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
US-09-917-384A-4

Query Match          79.8%; Score 71; DB 24; Length 153;
Best Local Similarity 100.0%; Pred. No. 2.5e-63;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 DNOIKPGLQLVNTGSSVDLSVTYVYWFTRDGGSSTLVYNCDDWAAMGCGNIRASFGSVN 77
Db 17 DNOIKPGLQLVNTGSSVDLSVTYVYWFTRDGGSSTLVYNCDDWAAMGCGNIRASFGSVN 76

Qy 78 PATPTADTYLQ 88
Db 77 PATPTADTYLQ 87

RESULT 14
US-09-917-384A-1
```

```
; Sequence 1, Application US/09917384A
; GENERAL INFORMATION:
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: DING, SHI-YOU
; APPLICANT: MCCARTER, SUZANNE
; APPLICANT: HIMMEL, MICHAEL E.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: VINZANT, TODD B.
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOYTICUS
; FILE REFERENCE: NREL 01-38
; CURRENT APPLICATION NUMBER: US/09/917,384A
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1121
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; US-09-917-384A-1

Query Match      79.8%; Score 71; DB 24; Length 1121;
Best Local Similarity 100.0%; Pred. No. 1.6e-62;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 DNQIKPGLQVNTGSSVDLSTVTYRYWFTRDGSSSTLVYNCDAWAAAGCGNIRASFGSVN 77
Db 51 DNQIKPGLQVNTGSSVDLSTVTYRYWFTRDGSSSTLVYNCDAWAAAGCGNIRASFGSVN 110

Qy 78 PATPTADTYLQ 88
Db 111 PATPTADTYLQ 121

RESULT 15
US-09-917-376-1
; Sequence 1, Application US/09917376
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOYTICUS
; FILE REFERENCE: 40197.4US01
; CURRENT APPLICATION NUMBER: US/09/917,376
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 957
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (957)
; OTHER INFORMATION: Any amino acid
; US-09-917-376-1

Query Match      40.4%; Score 36; DB 24; Length 957;
Best Local Similarity 100.0%; Pred. No. 4.9e-27;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 28 VNTGSSVDLSTVTYRYWFTRDGSSSTLVYNCDAWAA 63
Db 896 VNTGSSVDLSTVTYRYWFTRDGSSSTLVYNCDAWAA 931

RESULT 16
US-10-155-400-1
; Sequence 1, Application US/10155400
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOYTICUS
; FILE REFERENCE: NREL 01-36A
; CURRENT APPLICATION NUMBER: US/10/155,400
; CURRENT FILING DATE: 2002-10-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 957
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (957)
; OTHER INFORMATION: Any amino acid
; US-10-155-400-1

Query Match      40.4%; Score 36; DB 27; Length 957;
Best Local Similarity 100.0%; Pred. No. 4.9e-27;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 28 VNTGSSVDLSTVTYRYWFTRDGSSSTLVYNCDAWAA 63
Db 896 VNTGSSVDLSTVTYRYWFTRDGSSSTLVYNCDAWAA 931

Search completed: May 11, 2004, 12:32:34
Job time : 183 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 12:27:22 ; Search time 18 Seconds  
(without alignments)  
169.758 Million cell updates/sec

Title: US-09-917-376-4  
Perfect score: 89  
Sequence: 1 VSGGVKQYKRNDSAPGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 234425 seqs, 34333021 residues

Word size : 10

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Pending Patents AA New:\*  
1: /cgn2\_6/ptodata/2/paa/PCT\_NEW\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/paa/US06\_NEW\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/paa/US07\_NEW\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/paa/US08\_NEW\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/paa/US09\_NEW\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/paa/US10\_NEW\_COMB.pep:\*  
7: /cgn2\_6/ptodata/2/paa/US60\_NEW\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
---------------	-------	----------------	--------	----------	-------------

No matches found

Search completed: May 11, 2004, 12:33:04  
Job time : 18 secs

1: gb\_ba:\*  
2: gb\_ba:\*  
3: gb\_hg:\*  
4: gb\_hg:\*  
5: gb\_ov:\*  
6: gb\_ov:\*  
7: gb\_pat:\*  
8: gb\_pat:\*  
9: gb\_pl:\*  
10: gb\_pl:\*  
11: gb\_to:\*  
12: gb\_to:\*  
13: gb\_sy:\*  
14: gb\_sy:\*  
15: gb\_vl:\*  
16: gb\_vl:\*  
17: em\_fun:\*  
18: em\_fun:\*  
19: em\_hum:\*  
20: em\_hum:\*  
21: em\_in:\*  
22: em\_in:\*  
23: em\_mu:\*  
24: em\_mu:\*  
25: em\_or:\*  
26: em\_or:\*  
27: em\_ov:\*  
28: em\_ov:\*  
29: em\_pat:\*  
30: em\_pat:\*  
31: em\_ph:\*  
32: em\_ph:\*  
33: em\_pl:\*  
34: em\_pl:\*  
35: em\_ro:\*  
36: em\_ro:\*  
37: em\_sts:\*  
38: em\_sts:\*  
39: em\_un:\*  
40: em\_un:\*



```

Db      1483  ACGTGGCGGTACTGTTTACCCGGGATGGTGGTCTGCGACACTGGTGTAACACTGTGTGAC 1542
Qy      61   TrpAlaAlaMetGlyCySerGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db      1543  TGGGGGGGATGGGTGGGGAATATCCGCGCTCGTTCGGTGGTGAACCCGGCGAGC 1602
Qy      81   ProThrAlaAspThrTyrLeuGln 88
Db      1603  CCGACGGCGGACACCTACCTGCAG 1626

RESULT 2
AX700036
LOCUS      AX700036          3687 bp      DNA      linear      PAT 03-APR-2003
DEFINITION Sequence 2 from Patent WO03012095.
ACCESSION  AX700036
VERSION     AX700036.1 GI:29536019
KEYWORDS   .
SOURCE      synthetic construct
            synthetic construct
            artificial sequences.
REFERENCE   1
AUTHORS     Ding, S.Y., Adney, W.S., Vinzant, T.B., Himmel, M.E. and Decker, S.R.
TITLE       Thermal tolerant cellulase from Acidothermus cellulolyticus
JOURNAL     Patent: WO 0301209-A 2 13-FEB-2003;
            Midwest Research Institute (US)
FEATURES   source
            1..3687
            /organism="synthetic construct"
            /mol_type="unassigned DNA"
            /db_xref="taxon:32630"
            /note="Segment of Guxa"

ORIGIN
Alignment Scores:
Pred. No.:      3,236-78      Length:      3687
Score:          83.00        Matches:      83
Percent Similarity: 100.00%    Conservative: 0
Best Local Similarity: 100.00% Mismatches:    0
Query Match:     93.26%      Indels:       0
DB:              6          Gaps:         0

US-09-917-376-4 (1-89) x AX700036 (1-3687)

Qy      6   LysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysProGlyLeu 25
Db      1765  AAGTGCAGTACAGAACAAATGATTCGGCGCGGTGATACAGATCAACCGGGTCTC 1824
Qy      26   GlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArgTyrTrp 45
Db      1825  CAGTTGGTGAATACCGGGTCTGCTCGTGGATTGTCGACGGTACCGTGGTGGTCTGG 1884
Qy      46   PheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAlaMetGly 65
Db      1885  TTCACCCGGGATGGTGGTCTGCGACACTGGTGTACAACTGTGACTGGCGCGGATGGGG 1944
Qy      66   CysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAlaAspThr 85
Db      1945  TGTGGGAATATCCGGCCTCTGTTCCGCTCGGTGGTGAACCCGGCGACCGCGCGACACC 2004
Qy      86   TyrLeuGln 88
Db      2005  TACCTGCAG 2013

RESULT 3
AX700025
LOCUS      AX700025          3365 bp      DNA      linear      PAT 03-APR-2003
DEFINITION Sequence 2 from Patent WO03012095.
ACCESSION  AX700025
VERSION     AX700025.1 GI:29536018
KEYWORDS   .
SOURCE      Acidothermus cellulolyticus
            Acidothermus cellulolyticus

```

```

REFERENCE 1
AUTHORS     Adney, W.S., Ding, S.Y., Vinzant, T.B., Himmel, M.E., Decker, S.R. and
            Lantz McCarter, S.
TITLE       Thermal tolerant exoglucanase from Acidothermus cellulolyticus
JOURNAL     Patent: WO 03012095-A 2 13-FEB-2003;
            Midwest Research Institute (US)
FEATURES   source
            1..3365
            /organism="Acidothermus cellulolyticus"
            /mol_type="unassigned DNA"
            /db_xref="taxon:28049"

ORIGIN
Alignment Scores:
Pred. No.:      1,82e-65      Length:      3365
Score:          71.00        Matches:      71
Percent Similarity: 100.00%    Conservative: 0
Best Local Similarity: 100.00% Mismatches:    0
Query Match:     79.78%      Indels:       0
DB:              6          Gaps:         0

US-09-917-376-4 (1-89) x AX700025 (1-3365)

Qy      18   AspAsnGlnIleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeu 37
Db      151   GACAACCCAGATCAACACCGGGTCTCCAGTTGGTGAATACACGGGTCGTGGTGGATTG 210
Qy      38   SerThrValThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyr 57
Db      211   TCGACGGTGACGGTGGCGTACTGGTTTCAACCGGGATGGTGGTGGTGGTGGTGGTGGT 270
Qy      58   AsnCysAspTrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsn 77
Db      271   AACTGTGACTGGCGCGGATGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 330
Qy      78   ProAlaThrProThrAlaAspThrTyrLeuGln 88
Db      331   CCGGCGACGCCCGCGCGGACACCTACCTGCAG 363

RESULT 4
AX700058
LOCUS      AX700058          2869 bp      DNA      linear      PAT 03-APR-2003
DEFINITION Sequence 2 from Patent WO03012090.
ACCESSION  AX700058
VERSION     AX700058.1 GI:29536021
KEYWORDS   .
SOURCE      Acidothermus cellulolyticus
            Acidothermus cellulolyticus
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Frankineae; Acidothermaceae; Acidothermus.
REFERENCE   1
AUTHORS     Ding, S.Y., Adney, W.S., Vinzant, T.B. and Himmel, M.E.
TITLE       Thermal tolerant avicelase from Acidothermus cellulolyticus
JOURNAL     Patent: WO 03012090-A 2 13-FEB-2003;
            Midwest Research Institute (US)
FEATURES   source
            1..2869
            /organism="Acidothermus cellulolyticus"
            /mol_type="unassigned DNA"
            /db_xref="taxon:28049"

ORIGIN
Alignment Scores:
Pred. No.:      3,09e-28      Length:      2869
Score:          36.00        Matches:      36
Percent Similarity: 100.00%    Conservative: 0
Best Local Similarity: 100.00% Mismatches:    0
Query Match:     40.45%      Indels:       0
DB:              6          Gaps:         0

US-09-917-376-4 (1-89) x AX700058 (1-2869)

```

Qy	28	ValAsnThrGlySerSerSerValAspLeuSerThrValThrValArgTyrTrpPheThr	47
Db	2686	GTGAATACCGGGTCGTCGTCGGTGGATTGTGACGGTGACGGTCCGGTACTGGTTCACC	2745
Qy	48	ArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla	63
Db	2746	CGGGATGGTGGCTCGTCGACACTGGTGACAACTGTGACTGGGGGGCG	2793

Search completed: May 12, 2004, 16:05:36  
Job time : 2810 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 12, 2004, 11:39:06 ; Search time 322 Seconds  
(without alignments)  
1174.191 Million cell updates/sec

Title: US-09-917-376-4  
Perfect score: 89  
Sequence: 1 VSGGVQVQKNDSPGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table: OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Word size: 30

Total number of hits satisfying chosen parameters: 5

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2.1/USPTO.spool/US09917376/runat.11052004.114645.28650/app\_query.fasta.1.263  
-DB=N Geneseq 29Jan04 -QMT=fastap -SUFFIX=olip2n.rng -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=1000 -DOCALIGN=200 -THR SCORE=quality -THR MIN=30 -ALIGN=50 -MODE=LOCAL  
-OUTFMT=ptio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09917376@cgn.1.470@runat.11052004.114645.28650 -NCPUs=6 -ICPU=3  
-NO MMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_29Jan04.\*  
1: Geneseqn1980s.\*  
2: Geneseqn1990s.\*  
3: Geneseqn2000s.\*  
4: Geneseqn2001as.\*  
5: Geneseqn2001bs.\*  
6: Geneseqn2002s.\*  
7: Geneseqn2003as.\*  
8: Geneseqn2003bs.\*  
9: Geneseqn2003cs.\*  
10: Geneseqn2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	88	98.9	2289	7	Abz77633 Nucleotid
2	83	93.3	3687	7	Abz77634 Nucleotid
3	71	79.8	3365	7	Abz776162 A. cellul
4	36	40.4	2869	7	Abz77632 Nucleotid
5	36	40.4	2869	9	ADD22922 Acidothe

ALIGNMENTS

RESULT 1  
ABZ77633  
ID ABZ77633 standard; DNA; 2289 BP.  
XX  
AC ABZ77633;  
XX  
DT 03-JUN-2003 (first entry)  
XX  
DE Nucleotide sequence of the Mana polypeptide.  
XX  
KW Mana; mannanase A; glycoside hydrolase; enzyme; hemicellulose; sugar;  
KW food; feed; paper pulp; biofuel; mannanase; gene; ss.  
XX  
OS Acidothermus cellulolyticus.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..2289  
FT /\*tag= a  
FT /product= "Mana"  
XX  
PN WO2003012110-A1.  
XX  
PD 13-FEB-2003.  
XX  
PF 28-JUL-2001; 2001WO-US023819.  
XX  
PR 28-JUL-2001; 2001WO-US023819.  
XX  
PA (MIDE ) MIDWEST RES INST.  
XX  
PI Ding S, Adney WS, Vinzant TB, Himmel ME;  
XX  
DR WPI; 2003-248182/24.  
DR P-FSDB; ABP73022.  
XX  
PT Novel thermal tolerant mannanase A polypeptide derived from Acidothermus  
PT cellulolyticus, useful for reducing hemicellulose in a starting material,  
PT for processing of food, and as bulking agents in food stuffs.  
XX  
PS Example 1; Page 23; 46pp; English.  
XX  
CC The present sequence encodes Mana, a thermostable mannanase A polypeptide  
CC derived from Acidothermus cellulolyticus. Mana is a member of the  
CC glycoside hydrolase family of enzymes. Mana is useful for reducing  
CC hemicellulose in a starting material to simpler carbohydrate units, and  
CC ultimately to sugars which are useful in the food, feed, paper pulp, and  
CC biofuels industries. It is useful for the processing of food and in food  
CC stuffs as bulking agents, and for the degradation of mannan. Mana is  
CC also useful to raise polyclonal and monoclonal antibodies that are useful  
CC in purifying Mana, or detecting Mana polypeptide expression, and as well  
CC as reagent tools for characterizing the molecular actions of Mana  
CC polypeptides  
XX  
SQ Sequence 2289 BP; 463 A; 700 C; 688 G; 438 T; 0 U; 0 Other;  
  
Alignment Scores:  
Pred. No.: 2.32e-77 Length: 2289  
Score: 88.00 Matches: 88  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 98.88% Indels: 0  
DB: 7 Gaps: 0  
  
US-09-917-376-4 (1-89) x ABZ77633 (1-2289)  
  
Qy 1 ValserGlyGlyVallysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20  
Db 1363 GGTGCGGTGGGTGGAGGTGCGAGTACACACATGATTCGGCGCGGTGTATACCAG 1422  
Qy 21 lleyLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40  
Db 1423 ATCAAAACCGGGTCTCCAGTTGGTGAATACGGGGTCGTCTCGGTGGATTGTTCGACGGTG 1482

QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
 DB 1483 ACGGTGCGGTACTGTTCCACCGGGATGTTGGTCTGTCGACACTGTTACAACTGTGAC 1542

QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80  
 DB 1543 TGGGCGGCGATGGGTGTGGGAATATCCGGCTCGTTCGGTGAACCCGGCGACG 1602

QY 81 ProThrAlaAspThrTyrLeuGln 88  
 DB 1603 CCGACGGCGGACACTACTGCAG 1626

RESULT 2  
 ABZ77634  
 ID ABZ77634 standard; DNA; 3687 BP.

XX  
 AC ABZ77634;  
 XX  
 DT 03-JUN-2003 (first entry)  
 XX

DE Nucleotide sequence of the Guxa polypeptide.

XX Guxa; cellulase; glycoside hydrolase; enzyme; cellulose; sugar; biofuel;  
 KW detergent; pulp processing; paper processing; feed processing; textile;  
 KW gene; ss.

XX  
 OS Acidothermus cellulolyticus.

XX  
 FH Key Location/Qualifiers  
 FT CDS 1..3687  
 FT /\*tag= a  
 FT /product= "Guxa"

XX  
 PN WO2003012109-A1.  
 XX  
 PD 13-FEB-2003.

XX  
 PF 28-JUL-2001; 2001WO-US023817.  
 XX  
 PR 28-JUL-2001; 2001WO-US023817.  
 XX  
 PA (MIDE ) MIDWEST RES INST.

XX  
 PI Ding S, Adney WS, Vinzant TB, Himmel ME, Decker SR;  
 XX  
 DR WPI; 2003-239526/23.  
 DR P-PSDB; ABP73029.

XX  
 PT Novel thermal tolerant Guxa polypeptide derived from Acidothermus  
 PT cellulolyticus, useful for reducing cellulose in a starting material, and  
 PT for the conversion of biomass to biofuels and biofuel additives.

XX  
 PS Example 1; Page 23-24; 47pp; English.

XX  
 CC The present sequence encodes a Guxa polypeptide. Guxa is thermostable  
 CC cellulase, and is a member of the glycoside hydrolase family of enzymes.  
 CC Guxa is useful for reducing cellulose in a starting material such as  
 CC agricultural biomass to sugars. This is useful in biofuel production.  
 CC Guxa is also useful in the conversion of biomass to biofuels and biofuel  
 CC additives, in detergents, pulp and paper processing, food and feed  
 CC processing, and in textile processes. Guxa is also useful for raising  
 CC polyclonal and monoclonal antibodies that are useful in purifying Guxa,  
 CC or detecting Guxa polypeptide expression, as well as reagent tools for  
 CC characterizing the molecular actions of Guxa polypeptides

XX  
 SQ Sequence 3687 BP; 725 A; 1173 C; 1171 G; 618 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 3,238-72 Length: 3687  
 Score: 83.00 Matches: 83  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0

Query Match: 93.26% Indels: 0  
 DB: 7 Gaps: 0

US-09-917-376-4 (1-89) x ABZ77634 (1-3687)

QY 6 LysValGlnTyrLysAsnAsnSerAlaProGlyAspAsnClnIleLysProGlyLeu 25  
 DB 1765 AAGGTGCGAGTACAAGAACAAATGATTCGGCGCGGGTGATAACAGATCAAAACCGGGTCTC 1824

QY 26 GlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArgTyrTrp 45  
 DB 1825 CAGTTGGTGAATACCGGGTCTGTCGTGGATTGTCGACGGTGACGGTCCGTACTGG 1884

QY 46 PheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAlaMetGly 65  
 DB 1885 TTCACCGGGATCGTGGTCTGTCGACACTGTTGTTACAACTGTCACCTGGCGCGCATGGG 1944

QY 66 CysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAlaAspThr 85  
 DB 1945 TGTGGGAATATCCGGCTCGTTCGGTCCGTGAACCCGGCGACCGCGGACACC 2004

QY 86 TyrLeuGln 88  
 DB 2005 TACCTGCAG 2013

RESULT 3  
 ABZ76162  
 ID ABZ76162 standard; DNA; 3365 BP.

XX  
 AC ABZ76162;  
 XX  
 DT 29-MAY-2003 (first entry)  
 XX

DE A. cellulolyticus Gux1 protein encoding DNA.

XX  
 KW Gux1; thermal tolerant; exoglucanase; glycoside hydrolase; cellulose;  
 KW biofuel; detergent; pulp; paper processing; feed processing; textile;  
 KW cellulase; gene; ds.

XX  
 OS Acidothermus cellulolyticus.

XX  
 FH Key Location/Qualifiers  
 FT CDS 1..3365  
 FT /\*tag= a  
 FT /product= "Gux1"  
 FT /transl\_except= (pos: 682..683, aa: Pro)  
 FT /note= "this codon has an apparent one nucleotide  
 FT basepair deletion which alters the reading frame"

XX  
 PN WO2003012095-A1.  
 XX  
 PD 13-FEB-2003.

XX  
 PF 28-JUL-2001; 2001WO-US023820.

XX  
 PR 28-JUL-2001; 2001WO-US023820.

XX  
 PA (MIDE ) MIDWEST RES INST.

XX  
 PI Adney WS, Ding S, Vinzant TB, Himmel ME, Decker SR;  
 PI Lantz Mccarter S;  
 XX  
 DR WPI; 2003-300494/29.  
 DR P-PSDB; ABP71656.

XX  
 PT New thermal tolerant Gux1 peptide having specified amino acid sequence,  
 PT useful in the degradation of cellulose to biofuels.

XX  
 PS Disclosure; Page 22-23; 44pp; English.

XX  
 CC The invention relates to a thermal tolerant Gux1 peptide from A.  
 CC cellulolyticus. The Gux1 exoglucanase is a member of the glycoside  
 CC hydrolase family and comprises a catalytic domain GH48, carbohydrate

CC binding domain type III, and a carbohydrate binding domain type II. The  
CC polypeptide is useful in the degradation of cellulose into biofuel, or  
CC for conversion of biomass to biofuel additives. It is used in detergents,  
CC pulp and paper processing, food and feed processing, and in textile  
CC processing. It can also be used alone or in combination with other  
CC cellulase or glycoside hydrolases. The novel polypeptide generates  
CC alternative cellulase enzymes capable of assisting in the commercial-  
CC scale processing of cellulose to sugar for use in biofuel production. The  
CC present sequence represents a A. cellulolyticus Gux1 cellulase encoding  
CC DNA

XX Sequence 3365 BP; 669 A; 1117 C; 1041 G; 538 T; 0 U; 0 Other;

#### Alignment Scores:

Pred. No.: 2,29e-60 Length: 3365  
Score: 71.00 Matches: 71  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 79.78% Indels: 0  
DB: 7 Gaps: 0

US-09-917-376-4 (1-89) x ABZ76162 (1-3365)

QY 18 AsplandnlelvsProGlyLeuValenThrGlySerSerValAspLeu 37  
DB 151 GACAAACAGATCAACCGGGTCTCCAGTTGGTGAATACCGGGTCTCGTGGTGGATTG 210  
QY 38 SerThrValThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyr 57  
DB 211 TCGACGGTGAAGGCGGGTACTGGTTTACCCGGGATGGTGGTCTGCACACTGGTGAC 270  
QY 58 AsnCysAspTrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsn 77  
DB 271 AACTGTGACTGGGGGGGATGGGGTGTGGGATATCCGGCTCTGTTCCGGTCTGGTGAC 330

QY 78 ProIaThrProThrAlaAspThrTyrLeuGln 88

DB 331 CCGGCGAGCGCGGCGGACACCTACCTGCAG 363

#### RESULT 4

ID ABZ77632 standard; DNA; 2869 BP.

XX AC ABZ77632;

XX DT 03-JUN-2003 (first entry)

XX DE Nucleotide sequence of the avicelase Aviii.

XX KW Avicelase; Aviii; glycoside hydrolase; enzyme; cellulase; biofuel;  
XX detergent; pulp processing; paper processing; feed processing; textile;  
XX cellulose; gene; ss.

XX OS Acidothermus cellulolyticus.

XX FH Key Location/Qualifiers

XX CDS 1..2869 a

XX FT /tag= a

XX FT /product= "Aviii"

XX FT /transl\_except= (pos:2869,aa:Xaa)

XX FT /note= "Xaa is an unspecified residue"

XX PN WO2003012090-A2.

XX PD 13-FEB-2003.

XX PF 28-JUL-2001; 2001WO-US023818.

XX PR 28-JUL-2001; 2001WO-US023818.

XX PA (MIDE ) MIDWEST RES INST.

XX

PI Ding S, Adney WS, Vinzant TB, Himmel ME;

XX WPI; 2003-248177/24.

XX P-PSDB; ABP73015.

XX PT New thermostable Aviii peptide from Acidothermus cellulolyticus, useful  
XX for degradation of cellulose or in generating anti-Aviii antibodies for  
XX purifying recombinant Aviii polypeptides from genetically engineered  
XX host cells.

XX PS Claim 3; Page 24; 44pp; English.

XX CC The present sequence encodes a thermostable avicelase polypeptide,  
XX designated Aviii. Aviii is a member of the glycoside hydrolase family  
XX of enzymes, and is a cellulase. Aviii is useful in the conversion of  
XX biomass to biofuels and biofuel additives. It may be used in the  
XX production of detergents, pulp and paper processing, food and feed  
XX processing and in textile processes. The thermostable Aviii peptide is  
XX useful in the degradation of cellulose, and in generating specific anti-  
XX Aviii antibodies that are useful in purifying recombinant Aviii  
XX polypeptides from genetically engineered host cells, in detecting Aviii  
XX polypeptide expression, as well as a reagent tool for characterizing the  
XX molecular actions of the polypeptide. The Aviii polynucleotide is useful  
XX as a source of probes or primers in various diagnostic assays

XX SQ Sequence 2869 BP; 546 A; 857 C; 900 G; 565 T; 0 U; 1 Other;

#### Alignment Scores:

Pred. No.: 9,28e-26 Length: 2869  
Score: 36.00 Matches: 36  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 40.45% Indels: 0  
DB: 7 Gaps: 0

US-09-917-376-4 (1-89) x ABZ77632 (1-2869)

QY 28 ValAenThrGlySerSerSerValAspLeuSerThrValThrValArgTyrTrpPheThr 47

DB 2686 GTGAATACCGGGTCTCGTGGTGGATTGTGACGGTACGGTGGTACTGTTCCACC 2745

QY 48 ArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63

DB 2746 CGGGATGGTGGTCTCGTGCACACTGGTGTACAACTGTGACTGGGCGCG 2793

#### RESULT 5

ADD22922

ID ADD22922 standard; DNA; 2869 BP.

XX AC ADD22922;

XX DT 15-JAN-2004 (first entry)

XX DE Acidothermus cellulolyticus avicelase Aviii DNA.

XX KW Aviii; cellulose reduction; agricultural biomass; municipal solid waste;  
XX glycoside hydrolase; avicelase; ds; gene.

XX OS Acidothermus cellulolyticus.

XX FH Key Location/Qualifiers

XX CDS 1..2868 a

XX FT /tag= a

XX FT /product= "Aviii"

XX PN US2003108988-A1.

XX PD 12-JUN-2003.

XX PF 18-OCT-2002; 2002US-00155400.

XX PR 28-JUL-2001; 2001US-00917376.

XX

PA (DING/) DING S.  
 PA (ADNE/) ADNEY W S.  
 PA (VINZ/) VINZANT T B.  
 PA (HIMM/) HIMMEL M E.  
 XX  
 XX PI Ding S, Adney WS, Vinzant TB, Himmel ME;  
 XX  
 DR WPI; 2003-810853/76.  
 DR P-PSDB; ADD22921.  
 XX  
 PT New isolated thermal tolerant avicelase polynucleotide useful for  
 PT detection of a polynucleotide encoding AvilIII and for reducing cellulose  
 PT in a starting material, e.g. municipal solid waste.  
 XX  
 PS Claim 17; SEQ ID NO 2; 29pp; English.  
 XX  
 CC The invention relates to an isolated polynucleotide molecule encoding a  
 CC thermostable AvilIII polypeptide. The polynucleotide is useful for  
 CC detection of a polynucleotide encoding AvilIII. The polynucleotide is  
 CC useful for reducing cellulose in a starting material which involves  
 CC administering to the starting material, e.g. agricultural biomass or  
 CC municipal solid waste, a polypeptide molecule of the polynucleotide. The  
 CC method further comprises administering a second polypeptide molecule  
 CC chosen from the glycoside hydrolase family of proteins. The present  
 CC sequence represents DNA encoding the Acidothermus cellulolyticus  
 CC avicelase AvilIII.  
 XX  
 SQ Sequence 2869 BP; 546 A; 857 C; 900 G; 565 T; 0 U; 1 Other;

Alignment Scores:  
 Pred. No.: Length: 2869  
 Score: 36.00 Matches: 36  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 40.45% Indels: 0  
 DB: 9 Gaps: 0

US-09-917-376-4 (1-89) x ADD22922 (1-2869)  
 QY 28 ValAsnThrGlySerSerSerValAspLeuSerThrValThrValArgTyrTrpPheThr 47  
 Db 2686 GTGATACCGGGTCGTCGGTGGATTGTGACGGTGACGGTGGTACTGGTTACCC 2745  
 QY 48 ArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63  
 Db 2746 CGGGATGGTGGCTGTCGACACTGGGTGTACAACTGTGACTGGGGCGCG 2793

Search completed: May 12, 2004, 15:18:40  
 Job time : 324 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 12, 2004, 15:15:37 ; Search time 1965 Seconds  
(without alignments)  
1352.536 Million cell updates/sec

Title: US-09-917-376-4

Perfect score: 89

Sequence: 1 VSGGVKQYKNDSPGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table: OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Word size: 30

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Command line parameters:

-MODEL=frame\_p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO.spool/US09917376/runat\_11052004\_114646\_28669/app\_query.fasta\_1.263  
-DB=EST -QFMT=fastap -SUFFIX=olip2n.rst -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=oligo -TRANS=human40.cdi -LIST=1000  
-DOCALIGN=200 -THR SCORE=quality -THR MIN=30 -ALIGN=50 -MODE=LOCAL -OUTFMT=pto  
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09917376@cgn\_1\_1\_3437@runat\_11052004\_114646\_28669 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

EST:\*

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estmu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_htc:\*
- 9: gb\_est1:\*
- 10: gb\_est2:\*
- 11: gb\_hrc:\*
- 12: gb\_est3:\*
- 13: gb\_est4:\*
- 14: gb\_est5:\*
- 15: em\_estfun:\*
- 16: em\_estom:\*
- 17: em\_gss\_hum:\*
- 18: em\_gss\_inv:\*
- 19: em\_gss\_pln:\*
- 20: em\_gss\_vrt:\*
- 21: em\_gss\_fun:\*
- 22: em\_gss\_mam:\*
- 23: em\_gss\_mus:\*
- 24: em\_gss\_pro:\*
- 25: em\_gss\_rod:\*
- 26: em\_gss\_phg:\*
- 27: em\_gss\_vrl:\*
- 28: gb\_gssl:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

No matches found

Search completed: May 12, 2004, 16:38:30  
Job time : 1968 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 12, 2004, 15:18:47 ; Search time 67 Seconds  
(without alignments)  
737.174 Million cell updates/sec

Title: US-09-917-376-4  
Perfect score: 89  
Sequence: 1 VSGGVKQYKNNDSAPGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table: OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Word size: 30

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US09917376/runat\_11052004\_114646\_28682/app\_query.fasta\_1.263  
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=olip2n.rni -MINWATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=1000 -DOCALIGN=200 -THR SCORE=quality -THR MIN=30 -ALIGN=50 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09917376 @CGN 1 1.69 @runat\_11052004\_114646\_28682 -NCPU=6 -ICPU=3  
-NO\_WMAP -LARGQUERY -NEG\_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/pCTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
-----					

No matches found

Search completed: May 12, 2004, 16:39:50  
Job time : 69 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 12, 2004, 16:39:58 ; Search time 347 Seconds  
(without alignments)  
1161.871 Million cell updates/sec

Title: US-09-917-376-4  
Perfect score: 89  
Sequence: 1 VSGVKVQYKNDSPGDNQ.....RASFGVNPATPTADTYLQX 89

Scoring table:  
OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2941586 seqs, 2364995651 residues

Word size: 30

Total number of hits satisfying chosen parameters: 5

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Command line parameters:  
-MODEL=frame\_p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US09917376/runat\_11052004\_114648\_28756/app\_query.fasta\_1.263  
-DB=Published Applications NA -QFMT=fastap -SUFFIX=olip2n.rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=Oligo  
-TRANS=human40.cdi -LIST=1000 -DOALIGN=200 -THR SCORE=quality -THR\_MIN=30  
-ALIGN=50 -MODE=LOCAL -OUTFMT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US09917376 @CGN\_1\_1\_478 @runat\_11052004\_114648\_28756  
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60  
-XGAPOP=6 -XGAPEXT=7 -XGAPOP=60 -XGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Published Applications NA.\*  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/1/pubpna/US05\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq.\*  
17: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*  
18: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*  
19: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
------------	-------	-------	--------	----	----	-------------

1	88	98.9	2289	10	US-09-917-378-2	Sequence 2, Appli
2	83	93.3	3687	10	US-09-917-384-2	Sequence 2, Appli
3	83	93.3	3687	10	US-09-917-383-2	Sequence 2, Appli
4	36	40.4	2869	13	US-09-917-376-2	Sequence 2, Appli
5	36	40.4	2869	15	US-10-155-400-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1  
US-09-917-378-2  
; Sequence 2, Application US/09917378  
; Publication No. US20030119093A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: DECKER, STEPHEN R.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT MANNANASE FROM ACIDOTHERMUS  
; FILE REFERENCE: 40197-7US01  
; CURRENT APPLICATION NUMBER: US/09/917,378  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 2289  
; TYPE: DNA  
; ORGANISM: Acidothermus cellulolyticus  
US-09-917-378-2

Alignment Scores:  
Pred. No.: 4.3e-82 Length: 2289  
Score: 88.00 Matches: 88  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 98.88% Indels: 0  
DB: 10 Gaps: 0

US-09-917-376-4 (1-89) x US-09-917-378-2 (1-2289)

Qy	1	ValSerGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln	20
Db	1363	GHGTGGGTGGGTGAAGGTGAGTACAGACAAATGATTCGGCGCGGTGATAACAG	1422
Qy	21	IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal	40
Db	1423	ATCAACCGGGTCTCCAGTTGGTGAATACGGGGTCGTCGGTGGATTGTCGACGGTG	1482
Qy	41	ThrValArgTyrTrpPheThrArgAspGlySerSerThrLeuValTyrAsnCysAsp	60
Db	1483	ACGGTGGGTACTGGTTTACCCGGGATGGTGGTTCGTCGACACTGGTGTACACTGTGAC	1542
Qy	61	TpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr	80
Db	1543	TGGCGCGCATGGGGTGGGAATNTCCGCCCTCGTTCGGTGAACCCGGCGACG	1602
Qy	81	ProThrAlaAspThrTyrLeuGln	88
Db	1603	CCGACGGCGGACACCTACCTGCAG	1626

RESULT 2  
US-09-917-384-2  
; Sequence 2, Application US/09917384  
; Publication No. US20030096342A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: DECKER, STEPHEN R.  
; APPLICANT: HIMMEL, MICHAEL E.



RESULT 5

US-10-155-400-2  
 ; Sequence 2, Application US/10155400  
 ; Publication No. US20030108988A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: DING, SHI-YOU  
 ; APPLICANT: ADNEY, WILLIAM S.  
 ; APPLICANT: VINZANT, TODD B.  
 ; APPLICANT: HIMMEL, MICHAEL E.  
 ; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS  
 ; TITLE OF INVENTION: CELLULOLYTICUS  
 ; FILE REFERENCE: NREL 01-36A  
 ; CURRENT APPLICATION NUMBER: US/10/155,400  
 ; CURRENT FILING DATE: 2002-10-22  
 ; NUMBER OF SEQ ID NOS: 7  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 2  
 ; LENGTH: 2869  
 ; TYPE: DNA  
 ; ORGANISM: Acidothermus cellulolyticus  
 ; FEATURE:  
 ; NAME/KEY: modified\_base  
 ; LOCATION: (2869)  
 ; OTHER INFORMATION: a, c, t, g, other or unknown  
 US-10-155-400-2

Alignment Scores:  
 Pred. No.: 1,08e-27 Length: 2869  
 Score: 36.00 Matches: 36  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 40.45% Indels: 0  
 DB: 15 Gaps: 0

US-09-917-376-4 (1-89) x US-10-155-400-2 (1-2869)

QY	28	ValAsnThrGlySerSerValAspLeuSerThrValThrValArgTyrTrpPheThr	47
Db	2686	GTGATACCGGGTCGTCGGTGGATTGTCGACGGTGACGGTGGGTTCTCACC	2745
QY	48	ArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla	63
Db	2746	CGGGATGGTGGCTCGTCGACACTGGGTGTACAACTGTGACTGGGCGCG	2793

Search completed: May 12, 2004, 17:28:07  
 Job time : 349 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 11:39:45 ; Search time 14.4613 Seconds  
(without alignments)  
1738.900 Million cell updates/sec

Title: US-09-917-376-4

Perfect score: 470

Sequence: 1 VSGGVKQYKNDSPGDNQ.....RASFGVNPATPTADTYLQX 89

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

- 1: Geneseqp1980s:\*
- 2: Geneseqp1990s:\*
- 3: Geneseqp2000s:\*
- 4: Geneseqp2001s:\*
- 5: Geneseqp2002s:\*
- 6: Geneseqp2003as:\*
- 7: Geneseqp2003bs:\*
- 8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	469	99.8	88	7 ADD22925	Add22925 Acidother
2	469	99.8	89	7 ADD22924	Add22924 Acidother
3	469	99.8	154	6 ABP73020	Abp73020 Amino aci
4	469	99.8	762	6 ABP73022	Abp73022 Amino aci
5	466	99.1	150	6 ABP73025	Abp73025 Amino aci
6	466	99.1	1228	6 ABP73029	Abp73029 Amino aci
7	462	98.3	88	6 ABP73018	Abp73018 Amino aci
8	462	98.3	89	6 ABP73017	Abp73017 Amino aci
9	462	98.3	957	6 ABP73015	Abp73015 Amino aci
10	462	98.3	957	7 ADD22921	Add22921 Acidother
11	436	92.8	153	6 ABP71658	Abp71658 A. cellul
12	436	92.8	1121	6 ABP71656	Abp71656 A. cellul
13	231.5	49.3	782	2 AAR15625	Aar15625 Cellulase
14	195.5	41.6	616	2 AAV13494	AAV13494 Truncated
15	195.5	41.6	616	5 AAE16325	AAE16325 Active ce
16	194.5	41.4	1751	2 AAY13493	Aay13493 Truncated
17	194.5	41.4	1751	5 AAE16324	AAE16324 Active ce
18	191.5	40.7	1426	2 AAY13492	Aay13492 Truncated
19	191.5	40.7	1426	5 AAE16323	AAE16323 Active ce
20	188	40.0	477	5 AAO22444	AAO22444 Protein e
21	188	40.0	496	5 AAO22443	AAO22443 Protein e
22	188	40.0	677	5 AAU98063	Aau98063 Bacillus
23	186	39.6	499	2 AAR42122	Aar42122 NK-1 cell
24	172	36.6	700	2 AAR13227	Aar13227 Novel end
25	165.5	35.2	551	2 AAW18790	AAW18790 Corrected

26	165	35.1	167	2 AAR95080	Aar95080 Cellulose
27	165	35.1	476	3 AAY54123	Aay54123 A manana
28	165	35.1	493	2 AAY28850	Aay28850 Pectate 1
29	165	35.1	493	2 AAY43218	Aay43218 Pectate 1
30	165	35.1	531	2 AAW15238	Aaw15238 Scaffold 1
31	165	35.1	1853	2 AAW43108	Aaw43108 C. thermo
32	155.5	33.1	1352	4 AAG63962	Aag63962 Amino aci
33	150.5	32.0	1350	4 AAG63963	Aag63963 Amino aci
34	130.5	27.8	1483	6 ABU23559	Abu23559 Protein e
35	120.5	25.6	531	2 AAW01503	Aaw01503 60 kD end
36	120.5	25.6	532	2 AAR13229	Aar13229 Endogluca
37	114	24.3	162	2 AAR63634	Aar63634 Cellulose
38	114	24.3	162	2 AAW90077	Aaw90077 C. cellul
39	114	24.3	163	4 AAE05745	Aae05745 Clostridi
40	114	24.3	256	4 AAB81128	Aab81128 C17E2 Osp
41	114	24.3	256	5 AAU97869	Aau97869 E. coli c
42	114	24.3	328	4 AAE05749	Aae05749 Chimeric
43	114	24.3	341	4 AAE05747	Aae05747 Clostridi
44	114	24.3	428	4 AAE05748	Aae05748 Clostridi
45	113	24.0	190	4 AAE05746	Aae05746 Clostridi

#### ALIGNMENTS

#### RESULT 1

ADD22925

ID ADD22925 standard; protein; 88 AA.

XX AC ADD22925;

XX DT 15-JAN-2004 (first entry)

XX DE Acidothermus cellulolyticus avicelase Aviii CBD III #2.

XX enzyme; Aviii; cellulose reduction; agricultural biomass;

KW municipal solid waste; glycoside hydrolase; avicelase.

XX OS Acidothermus cellulolyticus.

XX PN US2003108988-A1.

XX PD 12-JUN-2003.

XX PF 18-OCT-2002; 2002US-00155400.

XX PR 28-JUL-2001; 2001US-00917376.

XX (DING/) DING S.

XX (ADNE/) ADNEY W S.

XX (VINZ/) VINZANT T B.

XX (HIMM/) HIMMEL M E.

PI Ding S, Adney WS, Vinzant TB, Himmel MB;

DR WPI; 2003-810853/76.

XX New isolated thermal tolerant avicelase polynucleotide useful for

PT detection of a polynucleotide encoding Aviii and for reducing cellulose

PT in a starting material, e.g. municipal solid waste.

XX Claim 16; SEQ ID NO 5; 29pp; English.

XX The invention relates to an isolated polynucleotide molecule encoding a

CC thermostable Aviii polypeptide. The polynucleotide is useful for

CC detection of a polynucleotide encoding Aviii. The polynucleotide is

CC useful for reducing cellulose in a starting material which involves

CC administering to the starting material, e.g. agricultural biomass or

CC municipal solid waste, a polypeptide molecule of the polynucleotide. The

CC method further comprises administering a second polypeptide molecule

CC chosen from the glycoside hydrolase family of proteins. The present

CC sequence represents the amino acid sequence of Acidothermus

CC cellulolyticus avicelase Aviii CBD III #2.

XX SQ Sequence 88 AA;  
 Query Match 99.8%; Score 469; DB 7; Length 88;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-46;  
 Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSTLVYNC 60  
 DB 1 VSGGVKQVQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSTLVYNC 60

QY 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88  
 DB 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88

RESULT 2  
 ADD22924  
 ID ADD22924 standard; protein; 89 AA.  
 XX AC ADD22924;  
 XX DT 15-JAN-2004 (first entry)  
 XX DE Acidothermus cellulolyticus avicelase AvIII CBD III #1.  
 XX KW enzyme; AvIII; cellulose reduction; agricultural biomass;  
 KW municipal solid waste; glycoside hydrolase; avicelase.  
 XX OS Acidothermus cellulolyticus.  
 XX FH Key Location/Qualifiers  
 FT Misc-difference 89 /label= Unknown  
 XX US2003108988-A1.  
 XX PN 12-JUN-2003.  
 XX PD 18-OCT-2002; 2002US-00155400.  
 XX PF 28-JUL-2001; 2001US-00917376.  
 XX PR (DING/) DING S.  
 PA (ADNE/) ADNEY W S.  
 PA (VINZ/) VINZANT T B.  
 PA (HIMM/) HIMMEL M E.  
 XX PI Ding S, Adney WS, Vinzant TB, Himmel ME;  
 XX WPI; 2003-810853/76.  
 XX DR New isolated thermal tolerant avicelase polynucleotide useful for  
 PT detection of a polynucleotide encoding AvIII and for reducing cellulose  
 PT in a starting material, e.g. municipal solid waste.  
 XX Claim 16; SEQ ID NO 4; 29pp; English.  
 XX CC The invention relates to an isolated polynucleotide molecule encoding a  
 CC thermostable AvIII polypeptide. The polynucleotide is useful for  
 CC detection of a polynucleotide encoding AvIII. The polynucleotide is  
 CC useful for reducing cellulose in a starting material which involves  
 CC administering to the starting material, e.g. agricultural biomass or  
 CC municipal solid waste, a polypeptide molecule of the polynucleotide. The  
 CC method further comprises administering a second polypeptide molecule  
 CC chosen from the glycoside hydrolase family of proteins. The present  
 CC sequence represents the amino acid sequence of Acidothermus  
 CC cellulolyticus avicelase AvIII CBD III #1.  
 XX SQ Sequence 89 AA;  
 Query Match 99.8%; Score 469; DB 7; Length 89;  
 Best Local Similarity 100.0%; Pred. No. 4e-46;

Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSTLVYNC 60  
 DB 1 VSGGVKQVQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSTLVYNC 60

QY 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88  
 DB 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88

RESULT 3  
 ABP73020  
 ID ABP73020 standard; peptide; 154 AA.  
 XX AC ABP73020;  
 XX DT 03-JUN-2003 (first entry)  
 XX DE Amino acid sequence of the ManA carbohydrate binding domain type III.  
 XX KW ManA; mannanase A; glycoside hydrolase; enzyme; hemicellulose; sugar;  
 KW food; feed; paper pulp; biofuel; mannanase.  
 XX OS Acidothermus cellulolyticus.  
 XX WO2003012110-A1.  
 XX PD 13-FEB-2003.  
 XX PF 28-JUL-2001; 2001WO-US023819.  
 XX PR 28-JUL-2001; 2001WO-US023819.  
 XX PA (WIDE ) MIDWEST RES INST.  
 XX PI Ding S, Adney WS, Vinzant TB, Himmel ME;  
 XX WPI; 2003-248182/24.  
 XX DR Novel thermal tolerant mannanase A polypeptide derived from Acidothermus  
 PT cellulolyticus, useful for reducing hemicellulose in a starting material,  
 PT for processing of food, and as bulking agents in food stuffs.  
 XX Claim 8; Page 7; 46pp; English.  
 XX CC The present sequence represents a fragment of ManA, a thermostable  
 CC mannanase A polypeptide derived from Acidothermus cellulolyticus. ManA is  
 CC a member of the glycoside hydrolase family of enzymes. ManA is useful for  
 CC reducing hemicellulose in a starting material to simpler carbohydrate  
 CC units, and ultimately to sugars which are useful in the food, feed, paper  
 CC pulp, and biofuels industries. It is useful for the processing of food  
 CC and in food stuffs as bulking agents, and for the degradation of mannanase.  
 CC ManA is also useful to raise polyclonal and monoclonal antibodies that  
 CC are useful in purifying ManA, or detecting ManA polypeptide expression,  
 CC and as well as reagent tools for characterizing the molecular actions of  
 CC ManA polypeptides  
 XX SQ Sequence 154 AA;  
 Query Match 99.8%; Score 469; DB 6; Length 154;  
 Best Local Similarity 100.0%; Pred. No. 7.8e-46;  
 Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSTLVYNC 60  
 DB 1 VSGGVKQVQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSTLVYNC 60

QY 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88  
 DB 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88

```

RESULT 4
ABP73022
ID ABP73022 standard; protein; 762 AA.
XX
AC ABP73022;
XX
DT 03-JUN-2003 (first entry)
XX
DE Amino acid sequence of the ManA polypeptide.
XX
KW ManA; mannanase A; glycoside hydrolase; enzyme; hemicellulose; sugar;
KW food; feed; paper pulp; biofuel; mannanase.
XX
OS Acidothermus cellulolyticus.
XX
PN WO2003012110-A1.
XX
PD 13-FEB-2003.
XX
PF 28-JUL-2001; 2001WO-US023819.
XX
PR 28-JUL-2001; 2001WO-US023819.
XX
PA (MIDE ) MIDWEST RES INST.
XX
PI Ding S, Adney WS, Vinzant TB, Himmel ME;
XX
DR WPI; 2003-248182/24.
DR N-PSDB; AB277633.
XX
XX
Novel thermal tolerant mannanase A polypeptide derived from Acidothermus
PT cellulolyticus, useful for reducing hemicellulose in a starting material,
PT for processing of food, and as bulking agents in food stuffs.
XX
PS Claim 3; Page 18-19; 46pp; English.
XX
XX
The present sequence represents ManA, a thermostable mannanase A
CC polypeptide derived from Acidothermus cellulolyticus. ManA is a member of
CC the glycoside hydrolase family of enzymes. ManA is useful for reducing
CC hemicellulose in a starting material to simpler carbohydrate units, and
CC ultimately to sugars which are useful in the food, feed, paper pulp, and
CC biofuels industries. It is useful for the processing of food and in food
CC stuffs as bulking agents, and for the degradation of mannanase. ManA is
CC also useful to raise polyclonal and monoclonal antibodies that are useful
CC in purifying ManA, or detecting ManA polypeptide expression, and as well
CC as reagent tools for characterizing the molecular actions of ManA
CC polypeptides
XX
SQ Sequence 762 AA;
XX
Query Match 99.8%; Score 469; DB 6; Length 762;
Best Local Similarity 100.0%; Pred. No. 5.7e-45;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VSGGVKVOYKNNDSAPGDNQIKFGLQLVNTGSSVDLSVTIVRYWFTRDGSSSTLVNCD 60
Db 455 VSGGVKVOYKNNDSAPGDNQIKFGLQLVNTGSSVDLSVTIVRYWFTRDGSSSTLVNCD 514
QY 61 WAAMGCNIRASFGSVNPATPTADTYLQ 88
Db 515 WAAMGCNIRASFGSVNPATPTADTYLQ 542
XX
RESULT 5
ABP73025
ID ABP73025 standard; peptide; 150 AA.
XX
AC ABP73025;
XX
DT 03-JUN-2003 (first entry)
XX
DE Amino acid sequence of the GuxA carbohydrate binding domain type III.
XX

```

```

KW GuxA; cellulase; glycoside hydrolase; enzyme; cellulose; sugar; biofuel;
KW detergent; pulp processing; paper processing; feed processing; textile.
XX
OS Acidothermus cellulolyticus.
XX
PN WO2003012109-A1.
XX
PD 13-FEB-2003.
XX
PF 28-JUL-2001; 2001WO-US023817.
XX
PR 28-JUL-2001; 2001WO-US023817.
XX
PA (MIDE ) MIDWEST RES INST.
XX
PI Ding S, Adney WS, Vinzant TB, Himmel ME, Decker SR;
XX
DR WPI; 2003-239526/23.
XX
XX
Novel thermal tolerant GuxA polypeptide derived from Acidothermus
PT cellulolyticus, useful for reducing cellulose in a starting material, and
PT for the conversion of biomass to biofuels and biofuel additives.
XX
PS Claim 6; Page 7; 47pp; English.
XX
XX
ABP73023-28 represent peptide fragments of a GuxA polypeptide. GuxA is a
CC thermostable cellulase, and is a member of the glycoside hydrolase family
CC of enzymes. GuxA is useful for reducing cellulose in a starting material
CC such as agricultural biomass to sugars. This is useful in biofuel
CC production. GuxA is also useful in the conversion of biomass to biofuels
CC and biofuel additives, in detergents, pulp and paper processing, food and
CC feed processing, and in textile process. GuxA is also useful for raising
CC polyclonal and monoclonal antibodies that are useful in purifying GuxA,
CC or detecting GuxA polypeptide expression, as well as reagent tools for
CC characterizing the molecular actions of GuxA polypeptides
XX
XX
SQ Sequence 150 AA;
XX
Query Match 99.1%; Score 466; DB 6; Length 150;
Best Local Similarity 98.9%; Pred. No. 1.7e-45;
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 VSGGVKVOYKNNDSAPGDNQIKFGLQLVNTGSSVDLSVTIVRYWFTRDGSSSTLVNCD 60
Db 1 VSGGLKVOYKNNDSAPGDNQIKFGLQLVNTGSSVDLSVTIVRYWFTRDGSSSTLVNCD 60
QY 61 WAAMGCNIRASFGSVNPATPTADTYLQ 88
Db 61 WAAMGCNIRASFGSVNPATPTADTYLQ 88
XX
RESULT 6
ABP73029
ID ABP73029 standard; protein; 1228 AA.
XX
AC ABP73029;
XX
DT 03-JUN-2003 (first entry)
XX
DE Amino acid sequence of the GuxA polypeptide.
XX
KW GuxA; cellulase; glycoside hydrolase; enzyme; cellulose; sugar; biofuel;
KW detergent; pulp processing; paper processing; feed processing; textile.
XX
OS Acidothermus cellulolyticus.
XX
PN WO2003012109-A1.
XX
PD 13-FEB-2003.
XX
PF 28-JUL-2001; 2001WO-US023817.
XX
PR 28-JUL-2001; 2001WO-US023817.

```

```

XX PA (MIDE ) MIDWEST RES INST.
XX PI Ding S, Adney WS, Vinzant TB, Himmel ME, Decker SR;
XX PI WPI; 2003-239526/23.
XX DR N-PSDB; ABZ77634.
XX DR Novel thermal tolerant GuxA polypeptide derived from Acidothermus
XX PT cellulolyticus, useful for reducing cellulose in a starting material, and
XX PT for the conversion of biomass to biofuels and biofuel additives.
XX XX
XX PS Claim 3; Page 19; 47pp; English.
XX CC The present sequence represents a GuxA polypeptide. GuxA is thermostable
XX CC cellulase, and is a member of the glycoside hydrolase family of enzymes.
XX CC GuxA is useful for reducing cellulose in a starting material such as
XX CC agricultural biomass to sugars. This is useful in biofuel production.
XX CC GuxA is also useful in the conversion of biomass to biofuels and biofuel
XX CC additives, in detergents, pulp and paper processing, food and feed
XX CC processing, and in textile processes. GuxA is also useful for raising
XX CC polyclonal and monoclonal antibodies that are useful in purifying GuxA,
XX CC or detecting GuxA polypeptide expression, as well as reagent tools for
XX CC characterizing the molecular actions of GuxA polypeptides
XX SQ Sequence 1228 AA;
Query Match 99.1%; Score 466; DB 6; Length 1228;
Best Local Similarity 98.9%; Pred. No. 2.3e-44;
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 VSGGVKVOYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRVYFTRDGGSSSTLVYVNC 60
DB 584 VSGGLKVOYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRVYFTRDGGSSSTLVYVNC 643
QY 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
DB 644 WAAMGCGNIRASFGSVNPATPTADTYLQ 671
RESULT 7
ABP73018
ID ABP73018 standard; peptide; 88 AA.
XX AC ABP73018;
XX AC ABP73018;
DT 03-JUN-2003 (first entry)
XX DE Amino acid sequence of AviiiI carbohydrate binding domain type II.
XX KW Avicelase; AviiiI; glycoside hydrolase; enzyme; cellulase; biofuel;
XX KW detergent; pulp processing; paper processing; feed processing; textile;
XX KW cellulose.
XX OS Acidothermus cellulolyticus.
XX FH Key Location/Qualifiers
FT Misc-difference 89
XX PN WO2003012090-A2.
XX PD 13-FEB-2003.
XX PF 28-JUL-2001; 2001WO-US023818.
XX PR 28-JUL-2001; 2001WO-US023818.
XX PA (MIDE ) MIDWEST RES INST.
XX PI Ding S, Adney WS, Vinzant TB, Himmel ME;
XX DR WPI; 2003-248177/24.
XX PT New thermostable AviiiI peptide from Acidothermus cellulolyticus, useful
XX PT for degradation of cellulose or in generating anti-AviiiI antibodies for
XX PT purifying recombinant AviiiI polypeptides from genetically engineered
XX PT host cells.
XX PS Claim 6; Page 8; 44pp; English.
XX XX

```

---

```

PT XX host cells.
XX PS Claim 6; Page 8; 44pp; English.
XX CC The present sequence is derived from a thermostable avicelase, designated
XX CC AviiiI. AviiiI is a member of the glycoside hydrolase family of enzymes,
XX CC and is a cellulase. AviiiI is useful in the conversion of biomass to
XX CC biofuels and biofuel additives. It may be useful in the production of
XX CC detergents, pulp and paper processing, food and feed processing and in
XX CC textile processes. The thermostable AviiiI peptide is useful in the
XX CC degradation of cellulose, and in generating specific anti-AviiiI
XX CC antibodies that are useful in purifying recombinant AviiiI polypeptides
XX CC from genetically engineered host cells, in detecting AviiiI polypeptide
XX CC expression, as well as a reagent tool for characterizing the molecular
XX CC actions of the polypeptide. The AviiiI polynucleotide is useful as a
XX CC source of probes or primers in various diagnostic assays
XX SQ Sequence 88 AA;
Query Match 98.3%; Score 462; DB 6; Length 88;
Best Local Similarity 97.7%; Pred. No. 2.5e-45;
Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 VSGGVKVOYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRVYFTRDGGSSSTLVYVNC 60
DB 1 VSGGVKVOYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRVYFTRDGGSSSTLVYVNC 60
QY 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
DB 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
RESULT 8
ABP73017
ID ABP73017 standard; peptide; 89 AA.
XX AC ABP73017;
XX DT 03-JUN-2003 (first entry)
XX DE Amino acid sequence of AviiiI carbohydrate binding domain type III.
XX KW Avicelase; AviiiI; glycoside hydrolase; enzyme; cellulase; biofuel;
XX KW detergent; pulp processing; paper processing; feed processing; textile;
XX KW cellulose.
XX OS Acidothermus cellulolyticus.
XX FH Key Location/Qualifiers
FT Misc-difference 89
XX PN WO2003012090-A2.
XX PD 13-FEB-2003.
XX PF 28-JUL-2001; 2001WO-US023818.
XX PR 28-JUL-2001; 2001WO-US023818.
XX PA (MIDE ) MIDWEST RES INST.
XX PI Ding S, Adney WS, Vinzant TB, Himmel ME;
XX DR WPI; 2003-248177/24.
XX PT New thermostable AviiiI peptide from Acidothermus cellulolyticus, useful
XX PT for degradation of cellulose or in generating anti-AviiiI antibodies for
XX PT purifying recombinant AviiiI polypeptides from genetically engineered
XX PT host cells.
XX PS Claim 6; Page 8; 44pp; English.
XX XX

```

CC The present sequence is derived from a thermostable avicelase, designated  
CC Aviii. Aviii is a member of the glycoside hydrolase family of enzymes,  
CC and is a cellulase. Aviii is useful in the conversion of biomass to  
CC biofuels and biofuel additives. It may be useful in the production of  
CC detergents, pulp and paper processing, food and feed processing and in  
CC textile processes. The thermostable Aviii peptide is useful in the  
CC degradation of cellulose, and in generating specific anti-Aviii  
CC antibodies that are useful in purifying recombinant Aviii polypeptides  
CC from genetically engineered host cells, in detecting Aviii polypeptide  
CC expression, as well as a reagent tool for characterizing the molecular  
CC actions of the polypeptide. The Aviii polynucleotide is useful as a  
CC source of probes or primers in various diagnostic assays  
XX

SQ Sequence 89 AA;  
Query Match 98.3%; Score 462; DB 6; Length 89;  
Best Local Similarity 97.7%; Pred. No. 2.5e-45;  
Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VSGGVKQYKNDSPAGDQIKPGLQVNTGSSVDLSTVTVRYWFTRDGSSSTLVNCD 60  
DB 1 VSGGVKQYKNDSPAGDQIKPGLQVNTGSSVDLSTVTVRYWFTRDGSSSTLVNCD 60  
QY 61 WAAMCGNIRASFGSVNPTADTYLQ 88  
DB 61 WAAIGCGNIRASFGSVNPTADTYLQ 88

RESULT 9  
ABP73015  
ID ABP73015 standard; protein; 957 AA.  
XX  
AC ABP73015;  
XX  
DT 03-JUN-2003 (first entry)  
XX  
DE Amino acid sequence of the avicelase Aviii.  
XX  
KW Avicelase; Aviii; glycoside hydrolase; enzyme; cellulase; biofuel;  
KW detergent; pulp processing; paper processing; feed processing; textile;  
KW cellulose.  
XX  
OS Acidothermus cellulolyticus.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 957 /note= "unspecified residue encoded by N"  
FT  
XX WO2003012090-A2.  
XX  
PD 13-FEB-2003.  
XX  
PF 28-JUL-2001; 2001WO-US023818.  
XX  
PR 28-JUL-2001; 2001WO-US023818.  
XX  
PA (MIDE ) MIDWEST RES INST.  
XX  
PI Ding S, Adney WS, Vinzant TB, Himmel ME;  
XX  
DR WPI; 2003-248177/24.  
DR N-PSDB; AB277632.  
XX  
PT New thermostable Aviii peptide from Acidothermus cellulolyticus, useful  
PT for degradation of cellulose or in generating anti-Aviii antibodies for  
PT purifying recombinant Aviii polypeptides from genetically engineered  
PT host cells.  
XX  
PS Claim 2; Page 20; 44pp; English.  
XX  
CC The present sequence represents a thermostable avicelase polypeptide,  
CC designated Aviii. Aviii is a member of the glycoside hydrolase family  
CC of enzymes, and is a cellulase. Aviii is useful in the conversion of

CC biomass to biofuels and biofuel additives. It may be useful in the  
CC production of detergents, pulp and paper processing, food and feed  
CC processing and in textile processes. The thermostable Aviii peptide is  
CC useful in the degradation of cellulose, and in generating specific anti-  
CC Aviii antibodies that are useful in purifying recombinant Aviii  
CC polypeptides from genetically engineered host cells, in detecting Aviii  
CC polypeptide expression, as well as a reagent tool for characterizing the  
CC molecular actions of the polypeptide. The Aviii polynucleotide is useful  
CC as a source of probes or primers in various diagnostic assays  
XX

SQ Sequence 957 AA;  
Query Match 98.3%; Score 462; DB 6; Length 957;  
Best Local Similarity 97.7%; Pred. No. 4.8e-44;  
Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VSGGVKQYKNDSPAGDQIKPGLQVNTGSSVDLSTVTVRYWFTRDGSSSTLVNCD 60  
DB 869 VSGGVKQYKNDSPAGDQIKPGLQVNTGSSVDLSTVTVRYWFTRDGSSSTLVNCD 928  
QY 61 WAAMCGNIRASFGSVNPTADTYLQ 88  
DB 929 WAAIGCGNIRASFGSVNPTADTYLQ 956

RESULT 10  
ADD22921  
ID ADD22921 standard; protein; 957 AA.  
XX  
AC ADD22921;  
XX  
DT 15-JAN-2004 (first entry)  
XX  
DE Acidothermus cellulolyticus avicelase Aviii.  
XX  
KW enzyme; Aviii; cellulose reduction; agricultural biomass;  
KW municipal solid waste; glycoside hydrolase; avicelase.  
XX  
OS Acidothermus cellulolyticus.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 957 /label= Unknown  
FT /note= "Encoded by N"  
XX  
PN US2003108988-A1.  
XX  
PD 12-JUN-2003.  
XX  
PF 18-OCT-2002; 2002US-00155400.  
XX  
PR 28-JUL-2001; 2001US-00917376.  
XX  
PA (DING/) DING S.  
PA (ADNE/) ADNEY W S.  
PA (VINZ/) VINZANT T B.  
PA (HIMM/) HIMMEL M E.  
XX  
PI Ding S, Adney WS, Vinzant TB, Himmel ME;  
XX  
DR WPI; 2003-810853/76.  
DR N-PSDB; ADD22922.  
XX  
PT New isolated thermal tolerant avicelase polynucleotide useful for  
PT detection of a polynucleotide encoding Aviii and for reducing cellulose  
PT in a starting material, e.g. municipal solid waste.  
XX  
PS Claim 16; SEQ ID NO 1; 29pp; English.  
XX  
CC The invention relates to an isolated polynucleotide molecule encoding a  
CC thermostable Aviii polypeptide. The polynucleotide is useful for  
CC detection of a polynucleotide encoding Aviii. The polynucleotide is  
CC useful for reducing cellulose in a starting material which involves



CC administering to the starting material, e.g. agricultural biomass or  
 CC municipal solid waste, a polypeptide molecule of the polynucleotide. The  
 CC method further comprises administering a second polypeptide molecule  
 CC chosen from the glycoside hydrolase family of proteins. The present  
 CC sequence represents the amino acid sequence of Acidothermus  
 CC cellulolyticus avicelase AvIII.  
 XX  
 SQ Sequence 957 AA;

Query Match 98.3%; Score 462; DB 7; Length 957;  
 Best Local Similarity 97.7%; Pred. No. 4.8e-44;  
 Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VSGGVKQVKNNDSPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGGSSTLVYNC 60  
 DB 869 VSGGVKQVKNNDSPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGGSSTLVYNC 928  
 QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
 DB 929 WAAMCGNIRASFGSVNPATPTADTYLQ 956

RESULT 11  
 ABP71658  
 ID ABP71658 standard; protein; 153 AA.  
 AC ABP71658;  
 XX  
 XX 29-MAY-2003 (first entry)  
 DE A. cellulolyticus Guxl protein CBD\_III domain fragment.  
 XX  
 XX Guxl; thermal tolerant; exoglucanase; glycoside hydrolase; cellulose;  
 KW biofuel; detergent; pulp; paper processing; feed processing; textile;  
 KW cellulase; enzyme.  
 XX  
 OS Acidothermus cellulolyticus.

PN W02003012095-A1.  
 XX 13-FEB-2003.  
 XX 28-JUL-2001; 2001WO-US023820.  
 XX 28-JUL-2001; 2001WO-US023820.  
 XX (WIDE ) MIDWEST RES INST.  
 XX Adney WS, Ding S, Vinzant TB, Himmel ME, Decker SR;  
 PI Lantz McCarter S;  
 XX  
 DR WPI; 2003-300494/29.  
 XX New thermal tolerant Guxl peptide having specified amino acid sequence,  
 PT useful in the degradation of cellulose to biofuels.  
 XX  
 PS Claim 2; Page 7; 44pp; English.

CC The invention relates to a thermal tolerant Guxl peptide from A.  
 CC cellulolyticus. The Guxl exoglucanase is a member of the glycoside  
 CC hydrolase family and comprises a catalytic domain GH48, carbohydrate  
 CC binding domain type III, and a carbohydrate binding domain type II. The  
 CC polypeptide is useful in the degradation of cellulose into biofuel, or  
 CC for conversion of biomass to biofuel additives. It is used in detergents,  
 CC pulp and paper processing, food and feed processing, and in textile  
 CC processing. It can also be used alone or in combination with other  
 CC cellulase or glycoside hydrolases. The novel polypeptide generates  
 CC alternative cellulase enzymes capable of assisting in the commercial-  
 CC scale processing of cellulose to sugar for use in biofuel production. The  
 CC present sequence represents a A. cellulolyticus Guxl cellulase CBD\_III  
 XX domain fragment  
 SQ Sequence 153 AA;

Query Match 92.8%; Score 436; DB 6; Length 153;  
 Best Local Similarity 96.4%; Pred. No. 4.9e-42;  
 Matches 81; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 5 VKVOYKNNDSPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGGSSTLVYNC 64  
 DB 4 LKQYKNNDSPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGGSSTLVYNC 63  
 QY 65 GCGNIRASFGSVNPATPTADTYLQ 88  
 DB 64 GCGNIRASFGSVNPATPTADTYLQ 87

RESULT 12  
 ABP71656  
 ID ABP71656 standard; protein; 1121 AA.  
 AC ABP71656;  
 XX  
 XX 29-MAY-2003 (first entry)  
 DE A. cellulolyticus Guxl protein.  
 XX  
 KW Guxl; thermal tolerant; exoglucanase; glycoside hydrolase; cellulose;  
 KW biofuel; detergent; pulp; paper processing; feed processing; textile;  
 KW cellulase; enzyme.  
 XX  
 OS Acidothermus cellulolyticus.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..34  
 FT /note= "potential signal peptide"  
 FT Protein 35..1121  
 FT /note= "mature protein"  
 FT Misc-difference 228  
 FT /note= "encoded by CG"  
 XX  
 PN W02003012095-A1.  
 XX 13-FEB-2003.  
 XX 28-JUL-2001; 2001WO-US023820.  
 XX 28-JUL-2001; 2001WO-US023820.  
 XX (WIDE ) MIDWEST RES INST.  
 XX Adney WS, Ding S, Vinzant TB, Himmel ME, Decker SR;  
 PI Lantz McCarter S;  
 XX  
 DR WPI; 2003-300494/29.  
 DR N-PSDB; ABZ76162.  
 XX New thermal tolerant Guxl peptide having specified amino acid sequence,  
 PT useful in the degradation of cellulose to biofuels.  
 XX  
 PS Claim 3; Page 18-19; 44pp; English.

CC The invention relates to a thermal tolerant Guxl peptide from A.  
 CC cellulolyticus. The Guxl exoglucanase is a member of the glycoside  
 CC hydrolase family and comprises a catalytic domain GH48, carbohydrate  
 CC binding domain type III, and a carbohydrate binding domain type II. The  
 CC polypeptide is useful in the degradation of cellulose into biofuel, or  
 CC for conversion of biomass to biofuel additives. It is used in detergents,  
 CC pulp and paper processing, food and feed processing, and in textile  
 CC processing. It can also be used alone or in combination with other  
 CC cellulase or glycoside hydrolases. The novel polypeptide generates  
 CC alternative cellulase enzymes capable of assisting in the commercial-  
 CC scale processing of cellulose to sugar for use in biofuel production. The  
 CC present sequence represents a A. cellulolyticus Guxl cellulase  
 XX Sequence 1121 AA;

ID	AA13494	standard; protein; 616 AA.
XX		
AC	AA13494;	
XX		
DT	30-JUL-1999	(first entry)
XX		
DE	Truncated cellulase Cel E3/B5.	
XX		
KW	Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;	
KW	Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;	
KW	cotton-containing fabric; stonewashing.	
XX		
OS	Unidentified.	
XX		
PN	EP921188-A2.	
XX		
PD	09-JUN-1999.	
XX		
PF	15-SEP-1998; 98EP-00810919.	
XX		
PR	19-SEP-1997; 97US-00932571.	
XX		
XX	(CLRN ) CLARIANT FINANCE BVI LTD.	
XX		
PI	Anderson P, Bergquist PL, Daniels RM, Farrington GK, Gibbs MD;	
PI	Morgan H, Williams P;	
XX		
DR	WPI; 1999-315403/27.	
XX		
DR	N-PSDB; AAX55660.	
XX		
PT	New truncated cellulase proteins, useful in detergents and for producing	
PT	'stonewashed' denim.	
XX		
PS	Claim 7; Page 42-43; 65pp; English.	
XX		
CC	The invention relates to a recombinant cellulase active protein free of	
CC	proteinases of native thermophilic and alkaliphilic origin, comprising	
CC	the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3,	
CC	Cel 6 or Cel E3/B5, or a stability region from one of the defined full-	
CC	length sequences, or functional equivalents. Cel B5 extends from amino	
CC	acid A1011 to P1424 or K1425 or N1426, and Cel B4/5 extends from amino	
CC	acid K635 to N1426 in the sequence shown in AA13492; Cel E1 extends from	
CC	amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel E1/2/3	
CC	extends from Y39 to G812, Cel E6 extends from amino acid V1233 to K1751	
CC	and the stability region extends from amino acid E482 to G635 in the	
CC	sequence shown in AA13493; Cel E3/B5 is shown in AA13494. The new	
CC	enzymes are useful in laundry detergent compositions to prevent or remove	
CC	staining, backtanning or graying, for use on cellulosic materials	
CC	including cotton-containing fabrics. They are especially useful for	
CC	preventing redeposition of colorant during stonewashing, and for	
CC	processing of textiles where cellulose breakdown is required. The new	
CC	truncated enzymes show reduced redeposition of dye compared to using non-	
CC	truncated cellulase compositions	
XX		
SQ	Sequence 616 AA;	
	Query Match	41.6%; Score 195.5; DB 2; Length 616;
	Best Local Similarity	43.2%; Pred. No. 1.3e-13;
	Matches	38; Conservative 15; Mismatches 34; Indels 1; Gaps 1;
QY	1	VSGGVKVVQYKNDNSAFGDNIQKPLQIVNTGSSVDLSTVTYVYWFTRDCGSSFLVYVNC
		D 60
DB	1	MGSQGVKLVYKNNFTSASTGSIRPWFKIVGGSSVDLSRVKIRYIVTDGDKPOSAV-CD
		59
QY	61	WAAMGCGNIRASFGSVNPTATPTADTYIQ 88
		: : : : :     : : : : :     : : : : :     : : : : :     : : : : :
DB	60	WAQIGASNTVTFNVKLSGSGVGDYYLE 87
		: : : : :     : : : : :     : : : : :     : : : : :     : : : : :
RESULT	15	
AAE16325		
ID	AAE16325	standard; protein; 616 AA.



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 11:39:45 ; Search time 14.2988 Seconds  
(without alignments)  
1738.900 Million cell updates/sec

Title: US-09-917-376-5

Perfect score: 469

Sequence: 1 VSGGVKQYKXNDSPAGDNQ.....IRASFGSVNPATPDYLYQ 88

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

- 1: geneseqp1980s:\*
- 2: geneseqp1990s:\*
- 3: geneseqp2000s:\*
- 4: geneseqp2001s:\*
- 5: geneseqp2002s:\*
- 6: geneseqp2003as:\*
- 7: geneseqp2003bs:\*
- 8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	469	100.0	88	7	ADD22925	Add22925 Acidother
2	469	100.0	89	7	ADD22924	Add22924 Acidother
3	469	100.0	154	6	ABP73020	ABP73020 Amino aci
4	469	100.0	762	6	ABP73022	ABP73022 Amino aci
5	466	99.4	150	6	ABP73025	ABP73025 Amino aci
6	466	99.4	1228	6	ABP73029	ABP73029 Amino aci
7	462	98.5	88	6	ABP73018	ABP73018 Amino aci
8	462	98.5	89	6	ABP73017	ABP73017 Amino aci
9	462	98.5	957	6	ABP73015	ABP73015 Amino aci
10	462	98.5	957	7	ADD22921	Add22921 Acidother
11	436	93.0	153	6	ABP71658	ABP71658 A. cellul
12	436	93.0	1121	6	ABP71656	ABP71656 A. cellul
13	231.5	49.4	782	2	AA15625	AA15625 Cellulase
14	195.5	41.7	616	2	AA13494	AA13494 Truncated
15	195.5	41.7	616	5	AA16325	AA16325 Active ce
16	194.5	41.5	1751	5	AA13493	AA13493 Truncated
17	194.5	41.5	1751	5	AA16324	AA16324 Active ce
18	191.5	40.8	1426	2	AA13492	AA13492 Truncated
19	191.5	40.8	1426	5	AA16323	AA16323 Active ce
20	188	40.1	477	5	AA022444	AA022444 Protein e
21	188	40.1	496	5	AA022443	AA022443 Protein e
22	188	40.1	677	5	AAU98063	AAU98063 Bacillus
23	186	39.7	499	2	AA42122	AA42122 NK-1 cell
24	172	36.7	700	2	AA13227	AA13227 Novel end
25	165.5	35.3	551	2	AAW18790	AAW18790 Corrected

26	165	35.2	167	2	AA95080	AA95080 Cellulose
27	165	35.2	476	3	AA54123	AA54123 A mamana
28	165	35.2	493	2	AA28850	AA28850 Pectate 1
29	165	35.2	493	2	AA43218	AA43218 Pectate 1
30	165	35.2	531	2	AAW15238	AAW15238 Scaffold
31	165	35.2	1853	2	AAW43108	AAW43108 C. thermo
32	155.5	33.2	1352	4	AA63962	AA63962 Amino aci
33	150.5	32.1	1350	4	AA63963	AA63963 Amino aci
34	130.5	27.8	1483	6	ABU23559	ABU23559 Protein e
35	120.5	25.7	531	2	AAW01503	AAW01503 60 kD end
36	120.5	25.7	532	2	AA13229	AA13229 Endogluca
37	114	24.3	162	2	AA96363	AA96363 Cellulose
38	114	24.3	162	2	AAW90077	AAW90077 C. cellul
39	114	24.3	163	4	AAE05745	AAE05745 Clostridi
40	114	24.3	256	4	AAE05745	AAE05745 Clostridi
41	114	24.3	256	5	AAU97869	AAU97869 E. coli c
42	114	24.3	328	4	AAE05749	AAE05749 Chimeric
43	114	24.3	341	4	AAE05747	AAE05747 Clostridi
44	114	24.3	428	4	AAE05748	AAE05748 Clostridi
45	113	24.1	190	4	AAE05746	AAE05746 Clostridi

ALIGNMENTS

RESULT 1  
ADD22925  
ID ADD22925 standard; protein; 88 AA.  
XX  
AC ADD22925;  
XX  
DT 15-JAN-2004 (first entry)  
XX  
DE Acidothermus cellulolyticus avicelase Aviii CBD III #2.  
XX  
KW enzyme; Aviii; cellulose reduction; agricultural biomass;  
KW municipal solid waste; glycoside hydrolase; avicelase.  
XX  
OS Acidothermus cellulolyticus.  
XX  
PN US2003108988-A1.  
XX  
PD 12-JUN-2003.  
XX  
PF 18-OCT-2002; 2002US-00155400.  
XX  
PR 28-JUL-2001; 2001US-00917376.  
XX  
PA (DING/) DING S.  
PA (ADNE/) ADNEY W S.  
PA (VINZ/) VINZANT T B.  
PA (HIMM/) HIMMEL M E.  
XX  
PI Ding S, Adney WS, Vinzant TB, Himmel MB;  
XX WPI; 2003-810853/76.  
XX  
PT New isolated thermal tolerant avicelase polynucleotide useful for  
PT detection of a polynucleotide encoding Aviii and for reducing cellulose  
PT in a starting material, e.g. municipal solid waste.  
XX  
PS Claim 16; SEQ ID NO 5; 29pp; English.  
XX  
CC The invention relates to an isolated polynucleotide molecule encoding a  
CC thermostable Aviii polypeptide. The polynucleotide is useful for  
CC detection of a polynucleotide encoding Aviii. The polynucleotide is  
CC useful for reducing cellulose in a starting material which involves  
CC administering to the starting material, e.g. agricultural biomass or  
CC municipal solid waste, a polypeptide molecule of the polynucleotide. The  
CC method further comprises administering a second polypeptide molecule  
CC chosen from the glycoside hydrolase family of proteins. The present  
CC sequence represents the amino acid sequence of Acidothermus  
CC cellulolyticus avicelase Aviii CBD III #2.

XX SQ Sequence 88 AA;

Query Match 100.0%; Score 469; DB 7; Length 88;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-46; Indels 0; Gaps 0;  
 Matches 88; Conservative 0; Mismatches 0;

QY 1 VSGGVKQVQKNNDSAPGDNQIKPGLQLVNTGSSVDLSVTYRYWFTRDGSSSTLVYVNC 60  
 |||||  
 Db 1 VSGGVKQVQKNNDSAPGDNQIKPGLQLVNTGSSVDLSVTYRYWFTRDGSSSTLVYVNC 60  
 |||||

QY 61 WAAMGCGNIRASFGSNPATPTADTYLQ 88  
 |||||  
 Db 61 WAAMGCGNIRASFGSNPATPTADTYLQ 88  
 |||||

RESULT 2  
 ADD22924  
 ID ADD22924 standard; protein; 89 AA.  
 XX  
 AC ADD22924;  
 XX  
 DT 15-JAN-2004 (first entry)  
 XX  
 DE Acidotherrnus cellulolyticus avicelase AvIII CBD III #1.  
 XX  
 KW enzyme; AvIII; cellulose reduction; agricultural biomass;  
 KW municipal solid waste; glycoside hydrolase; avicelase.  
 XX  
 OS Acidotherrnus cellulolyticus.  
 XX  
 Key Location/Qualifiers  
 FH Misc-difference 89 /label= Unknown  
 FT  
 FT  
 XX US2003108988-A1.  
 PN  
 XX 12-JUN-2003.  
 PD  
 XX 18-OCT-2002; 2002US-00155400.  
 XX  
 PF 28-JUL-2001; 2001US-00917376.  
 PR  
 XX (DING/) DING S.  
 PA (ADNE/) ADNEY W S.  
 PA (VINZ/) VINZANT T B.  
 PA (HIMM/) HIMMEL M E.  
 XX  
 PI Ding S, Adney WS, Vinzant TB, Himmel ME;  
 XX WPI; 2003-810853/76.  
 DR  
 XX New isolated thermal tolerant avicelase polynucleotide useful for  
 PT detection of a polynucleotide encoding AvIII and for reducing cellulose  
 PT in a starting material, e.g. municipal solid waste.  
 XX  
 PS Claim 15; SEQ ID NO 4; 29pp; English.  
 XX  
 CC The invention relates to an isolated polynucleotide molecule encoding a  
 CC thermostable AvIII polypeptide. The polynucleotide is useful for  
 CC detection of a polynucleotide encoding AvIII. The polynucleotide is  
 CC useful for reducing cellulose in a starting material which involves  
 CC administering to the starting material, e.g. agricultural biomass or  
 CC municipal solid waste, a polypeptide molecule of the polynucleotide. The  
 CC method further comprises administering a second polypeptide molecule  
 CC chosen from the glycoside hydrolase family of proteins. The present  
 CC sequence represents the amino acid sequence of Acidotherrnus  
 CC cellulolyticus avicelase AvIII CBD III #1.  
 XX  
 SQ Sequence 89 AA;

Query Match 100.0%; Score 469; DB 7; Length 89;  
 Best Local Similarity 100.0%; Pred. No. 4e-46;

Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVQKNNDSAPGDNQIKPGLQLVNTGSSVDLSVTYRYWFTRDGSSSTLVYVNC 60  
 |||||  
 Db 1 VSGGVKQVQKNNDSAPGDNQIKPGLQLVNTGSSVDLSVTYRYWFTRDGSSSTLVYVNC 60  
 |||||

QY 61 WAAMGCGNIRASFGSNPATPTADTYLQ 88  
 |||||  
 Db 61 WAAMGCGNIRASFGSNPATPTADTYLQ 88  
 |||||

RESULT 3  
 ABP73020  
 ID ABP73020 standard; peptide; 154 AA.  
 XX  
 AC ABP73020;  
 XX  
 DT 03-JUN-2003 (first entry)  
 XX  
 DE Amino acid sequence of the ManA carbohydrate binding domain type III.  
 XX  
 KW ManA; mannanase A; glycoside hydrolase; enzyme; hemicellulose; sugar;  
 KW food; feed; paper pulp; biofuel; mannanase.  
 XX  
 OS Acidotherrnus cellulolyticus.  
 XX  
 PN WO2003012110-A1.  
 XX  
 PD 13-FEB-2003.  
 XX  
 PF 28-JUL-2001; 2001WO-US023819.  
 XX  
 PR 28-JUL-2001; 2001WO-US023819.  
 XX  
 PA (WIDE ) MIDWEST RES INST.  
 XX  
 PI Ding S, Adney WS, Vinzant TB, Himmel ME;  
 XX WPI; 2003-248182/24.  
 DR  
 XX Novel thermal tolerant mannanase A polypeptide derived from Acidotherrnus  
 PT cellulolyticus, useful for reducing hemicellulose in a starting material,  
 PT for processing of food, and as bulking agents in food stuffs.  
 XX  
 PS Claim 8; Page 7; 46pp; English.  
 XX  
 CC The present sequence represents a fragment of ManA, a thermostable  
 CC mannanase A polypeptide derived from Acidotherrnus cellulolyticus. ManA is  
 CC a member of the glycoside hydrolase family of enzymes. ManA is useful for  
 CC reducing hemicellulose in a starting material to simpler carbohydrate  
 CC units, and ultimately to sugars which are useful in the food, feed, paper  
 CC pulp, and biofuels industries. It is useful for the processing of food  
 CC and in food stuffs as bulking agents, and for the degradation of mannanase.  
 CC ManA is also useful to raise polyclonal and monoclonal antibodies that  
 CC are useful in purifying ManA, or detecting ManA polypeptide expression,  
 CC and as well as reagent tools for characterizing the molecular actions of  
 CC ManA polypeptides  
 XX  
 SQ Sequence 154 AA;

Query Match 100.0%; Score 469; DB 6; Length 154;  
 Best Local Similarity 100.0%; Pred. No. 7.8e-46;  
 Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVQKNNDSAPGDNQIKPGLQLVNTGSSVDLSVTYRYWFTRDGSSSTLVYVNC 60  
 |||||  
 Db 1 VSGGVKQVQKNNDSAPGDNQIKPGLQLVNTGSSVDLSVTYRYWFTRDGSSSTLVYVNC 60  
 |||||

QY 61 WAAMGCGNIRASFGSNPATPTADTYLQ 88  
 |||||  
 Db 61 WAAMGCGNIRASFGSNPATPTADTYLQ 88  
 |||||

RESULT 4  
 ABP73022  
 ID ABP73022 standard; protein; 762 AA.  
 XX AC ABP73022;  
 XX DT 03-JUN-2003 (first entry)  
 XX DE Amino acid sequence of the ManA polypeptide.  
 XX KW ManA; mannanase A; glycoside hydrolase; enzyme; hemicellulose; sugar;  
 XX KW food; feed; paper pulp; biofuel; mannanase.  
 XX OS Acidotherrmus cellulolyticus.  
 XX PN WO2003012110-A1.  
 XX PD 13-FEB-2003.  
 XX PF 28-JUL-2001; 2001WO-US023819.  
 XX PR 28-JUL-2001; 2001WO-US023819.  
 XX PA (MIDE ) MIDWEST RES INST.  
 XX PI Ding S, Adney WS, Vinzant TB, Himmel ME;  
 XX DR WPI; 2003-248182/24.  
 XX DR N-PSDB; ABZ77633.  
 XX PT Novel thermal tolerant mannanase A polypeptide derived from Acidotherrmus  
 PT cellulolyticus, useful for reducing hemicellulose in a starting material,  
 PT for processing of food, and as bulking agents in food stuffs.  
 XX PS Claim 3; Page 18-19; 46pp; English.  
 XX CC The present sequence represents ManA, a thermostable mannanase A  
 CC polypeptide derived from Acidotherrmus cellulolyticus. ManA is a member of  
 CC the glycoside hydrolase family of enzymes. ManA is useful for reducing  
 CC hemicellulose in a starting material to simpler carbohydrate units, and  
 CC ultimately to sugars which are useful in the food, feed, paper pulp, and  
 CC biofuels industries. It is useful for the processing of food and in food  
 CC stuffs as bulking agents, and for the degradation of mannanase. ManA is  
 CC also useful to raise polyclonal and monoclonal antibodies that are useful  
 CC in purifying ManA, or detecting ManA polypeptide expression, and as well  
 CC as reagent tools for characterizing the molecular actions of ManA  
 CC polypeptides  
 XX SQ Sequence 762 AA;  
 Query Match 100.0%; Score 469; DB 6; Length 762;  
 Best Local Similarity 100.0%; Pred. No. 5.7e-45;  
 Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VSGGVKQVQKNDSPAGDNQIKPGLQLVNTGSSVDLSVTIVRYWFTTRDGGSSSTLVYNC 60  
 DB 455 VSGGVKQVQKNDSPAGDNQIKPGLQLVNTGSSVDLSVTIVRYWFTTRDGGSSSTLVYNC 514  
 QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
 DB 515 WAAMCGNIRASFGSVNPATPTADTYLQ 542  
 RESULT 5  
 ABP73025  
 ID ABP73025 standard; peptide; 150 AA.  
 XX AC ABP73025;  
 XX DT 03-JUN-2003 (first entry)  
 XX DE Amino acid sequence of the GuxA carbohydrate binding domain type III.

KW GuxA; cellulase; glycoside hydrolase; enzyme; cellulose; sugar; biofuel;  
 KW detergent; pulp processing; paper processing; feed processing; textile.  
 XX OS Acidotherrmus cellulolyticus.  
 XX PN WO2003012109-A1.  
 XX PD 13-FEB-2003.  
 XX PF 28-JUL-2001; 2001WO-US023817.  
 XX PR 28-JUL-2001; 2001WO-US023817.  
 XX PA (MIDE ) MIDWEST RES INST.  
 XX PI Ding S, Adney WS, Vinzant TB, Himmel ME, Decker SR;  
 XX DR WPI; 2003-239526/23.  
 XX PT Novel thermal tolerant GuxA polypeptide derived from Acidotherrmus  
 PT cellulolyticus, useful for reducing cellulose in a starting material, and  
 PT for the conversion of biomass to biofuels and biofuel additives.  
 XX PS Claim 6; Page 7; 47pp; English.  
 XX CC ABP73023-28 represent peptide fragments of a GuxA polypeptide. GuxA is a  
 CC thermostable cellulase, and is a member of the glycoside hydrolase family  
 CC of enzymes. GuxA is useful for reducing cellulose in a starting material  
 CC such as agricultural biomass to sugars. This is useful in biofuel  
 CC production. GuxA is also useful in the conversion of biomass to biofuels  
 CC and biofuel additives, in detergents, pulp and paper processing, food and  
 CC feed processing, and in textile process. GuxA is also useful for raising  
 CC polyclonal and monoclonal antibodies that are useful in purifying GuxA,  
 CC or detecting GuxA polypeptide expression, as well as reagent tools for  
 CC characterizing the molecular actions of GuxA polypeptides  
 XX SQ Sequence 150 AA;  
 Query Match 99.4%; Score 466; DB 6; Length 150;  
 Best Local Similarity 98.9%; Pred. No. 1.7e-45;  
 Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VSGGVKQVQKNDSPAGDNQIKPGLQLVNTGSSVDLSVTIVRYWFTTRDGGSSSTLVYNC 60  
 DB 1 VSGGLKQVQKNDSPAGDNQIKPGLQLVNTGSSVDLSVTIVRYWFTTRDGGSSSTLVYNC 60  
 QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
 DB 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
 RESULT 6  
 ABP73029  
 ID ABP73029 standard; protein; 1228 AA.  
 XX AC ABP73029;  
 XX DT 03-JUN-2003 (first entry)  
 XX DE Amino acid sequence of the GuxA polypeptide.  
 XX KW GuxA; cellulase; glycoside hydrolase; enzyme; cellulose; sugar; biofuel;  
 KW detergent; pulp processing; paper processing; feed processing; textile.  
 XX OS Acidotherrmus cellulolyticus.  
 XX PN WO2003012109-A1.  
 XX PD 13-FEB-2003.  
 XX PF 28-JUL-2001; 2001WO-US023817.  
 XX PR 28-JUL-2001; 2001WO-US023817.

```
XX PA (MIDE ) MIDWEST RES INST.
XX PI Ding S, Adney WS, Vinzant TB, Himmel ME, Decker SR;
XX XX
XX DR WPI; 2003-239526/23.
XX N-PSDB; ABZ77634.
XX XX
XX PT Novel thermal tolerant GuxA polypeptide derived from Acidothermus
XX cellulolyticus, useful for reducing cellulose in a starting material, and
XX for the conversion of biomass to biofuels and biofuel additives.
XX PS Claim 3; Page 19; 47pp; English.
XX XX
XX CC The present sequence represents a GuxA polypeptide. GuxA is thermostable
XX cellulase, and is a member of the glycoside hydrolase family of enzymes.
XX CC GuxA is useful for reducing cellulose in a starting material such as
XX agricultural biomass to sugars. This is useful in biofuel production.
XX CC GuxA is also useful in the conversion of biomass to biofuels and biofuel
XX additives, in detergents, pulp and paper processing, food and feed
XX processing, and in textile processes. GuxA is also useful for raising
XX polyclonal and monoclonal antibodies that are useful in purifying GuxA,
XX or detecting GuxA polypeptide expression, as well as reagent tools for
XX characterizing the molecular actions of GuxA polypeptides
XX SQ Sequence 1228 AA;
Query Match 99.4%; Score 466; DB 6; Length 1228;
Best Local Similarity 98.9%; Pred. No. 2.3e-44; Indels 0; Gaps 0;
Matches 87; Conservative 1; Mismatches 0;
QY 1 VSGGVKQVQKNNDSAPGDNQIKPGLQLVNTGSSVDLSVTYRVYFTRDGGSSTLVYVNC 60
DB 584 VSGGLKVQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSVTYRVYFTRDGGSSTLVYVNC 643
QY 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
DB 644 WAAMGCGNIRASFGSVNPATPTADTYLQ 671
RESULT 7
ABP73018
ID ABP73018 standard; peptide; 88 AA.
XX AC ABP73018;
XX DT 03-JUN-2003 (first entry)
XX DE Amino acid sequence of AvIII carbohydrate binding domain type II.
XX KW Avicelase; AvIII; glycoside hydrolase; enzyme; cellulase; biofuel;
XX KW detergent; pulp processing; paper processing; feed processing; textile;
XX KW cellulose.
XX OS Acidothermus cellulolyticus.
XX FH Key Location/Qualifiers
XX FT Misc-difference 89 /note= "unspecified residue"
XX XX
XX PN WO2003012090-A2.
XX XX
XX PD 13-FEB-2003.
XX XX
XX PF 28-JUL-2001; 2001WO-US023818.
XX XX
XX PR 28-JUL-2001; 2001WO-US023818.
XX PA (MIDE ) MIDWEST RES INST.
XX PI Ding S, Adney WS, Vinzant TB, Himmel ME;
XX WPI; 2003-248177/24.
XX XX
XX PT New thermostable AvIII peptide from Acidothermus cellulolyticus, useful
XX for degradation of cellulose or in generating anti-AvIII antibodies for
XX purifying recombinant AvIII polypeptides from genetically engineered
XX T
```

```
PT host cells.
XX Claim 6; Page 8; 44pp; English.
XX CC The present sequence is derived from a thermostable avicelase, designated
XX AvIII. AvIII is a member of the glycoside hydrolase family of enzymes,
XX and is a cellulase. AvIII is useful in the conversion of biomass to
XX biofuels and biofuel additives. It may be useful in the production of
XX detergents, pulp and paper processing, food and feed processing and in
XX textile processes. The thermostable AvIII peptide is useful in the
XX degradation of cellulose, and in generating specific anti-AvIII
XX antibodies that are useful in purifying recombinant AvIII polypeptides
XX from genetically engineered host cells, in detecting AvIII polypeptide
XX expression, as well as a reagent tool for characterizing the molecular
XX actions of the polypeptide. The AvIII polynucleotide is useful as a
XX source of probes or primers in various diagnostic assays
XX SQ Sequence 88 AA;
Query Match 98.5%; Score 462; DB 6; Length 88;
Best Local Similarity 97.7%; Pred. No. 2.5e-45;
Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 VSGGVKQVQKNNDSAPGDNQIKPGLQLVNTGSSVDLSVTYRVYFTRDGGSSTLVYVNC 60
DB 1 VSGGVKQVQKNNDSAPGDNQIKPGLQLVNTGSSVDLSVTYRVYFTRDGGSSTLVYVNC 60
QY 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
DB 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
RESULT 8
ABP73017
ID ABP73017 standard; peptide; 89 AA.
XX AC ABP73017;
XX DT 03-JUN-2003 (first entry)
XX DE Amino acid sequence of AvIII carbohydrate binding domain type III.
XX KW Avicelase; AvIII; glycoside hydrolase; enzyme; cellulase; biofuel;
XX KW detergent; pulp processing; paper processing; feed processing; textile;
XX KW cellulose.
XX OS Acidothermus cellulolyticus.
XX FH Key Location/Qualifiers
XX FT Misc-difference 89 /note= "unspecified residue"
XX XX
XX PN WO2003012090-A2.
XX XX
XX PD 13-FEB-2003.
XX XX
XX PF 28-JUL-2001; 2001WO-US023818.
XX XX
XX PR 28-JUL-2001; 2001WO-US023818.
XX PA (MIDE ) MIDWEST RES INST.
XX PI Ding S, Adney WS, Vinzant TB, Himmel ME;
XX WPI; 2003-248177/24.
XX XX
XX PT New thermostable AvIII peptide from Acidothermus cellulolyticus, useful
XX for degradation of cellulose or in generating anti-AvIII antibodies for
XX purifying recombinant AvIII polypeptides from genetically engineered
XX PT host cells.
XX PS Claim 6; Page 8; 44pp; English.
XX XX
```

CC The present sequence is derived from a thermostable avicelase, designated  
CC Aviii. Aviii is a member of the glycoside hydrolase family of enzymes,  
CC and is a cellulase. Aviii is useful in the conversion of biomass to  
CC biofuels and biofuel additives. It may be useful in the production of  
CC detergents, pulp and paper processing, food and feed processing and in  
CC textile processes. The thermostable Aviii peptide is useful in the  
CC degradation of cellulose, and in generating specific anti-Aviii  
CC antibodies that are useful in purifying recombinant Aviii polypeptides  
CC from genetically engineered host cells, in detecting Aviii polypeptide  
CC expression, as well as a reagent tool for characterizing the molecular  
CC actions of the polypeptide. The Aviii polynucleotide is useful as a  
CC source of probes or primers in various diagnostic assays  
XX  
SQ Sequence 89 AA;

Query Match 98.5%; Score 462; DB 6; Length 89;  
Best Local Similarity 97.7%; Pred. No. 2.5e-45;  
Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VSGGVKQYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTYVWFTRDGSSSTLVYVNC 60  
DB 1 VSGGVKQYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTYVWFTRDGSSSTLVYVNC 60  
QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
DB 61 WAAIGCGNIRASFGSVNPATPTADTYLQ 88

RESULT 9  
ABP73015  
ID ABP73015 standard; protein; 957 AA.  
AC ABP73015;  
XX  
XX 03-JUN-2003 (first entry)  
DE Amino acid sequence of the avicelase Aviii.  
KW Avicelase; Aviii; glycoside hydrolase; enzyme; cellulase; biofuel;  
KW detergent; pulp processing; paper processing; feed processing; textile;  
KW cellulose.

OS Acidothermus cellulolyticus.  
FH Key Location/Qualifiers  
FT Misc-difference 957 /note= "unspecified residue encoded by N"  
FT  
XX  
XX WO2003012090-A2.  
XX  
XX 13-FEB-2003.  
XX  
XX 28-JUL-2001; 2001WO-US023818.  
XX  
XX 28-JUL-2001; 2001WO-US023818.  
XX (MIDE ) MIDWEST RES INST.  
XX

PI Ding S, Adney WS, Vinzant TB, Himmel ME;  
XX  
XX WPI; 2003-248177/24.  
XX N-PSDB; ABZ77632.  
XX  
XX New thermostable Aviii peptide from Acidothermus cellulolyticus, useful  
XX for degradation of cellulose or in generating anti-Aviii antibodies for  
XX purifying recombinant Aviii polypeptides from genetically engineered  
XX host cells.  
XX  
XX Claim 2; Page 20; 44pp; English.

CC The present sequence represents a thermostable avicelase polypeptide,  
CC designated Aviii. Aviii is a member of the glycoside hydrolase family  
CC of enzymes, and is a cellulase. Aviii is useful in the conversion of  
XX

CC biomass to biofuels and biofuel additives. It may be useful in the  
CC production of detergents, pulp and paper processing, food and feed  
CC processing and in textile processes. The thermostable Aviii peptide is  
CC useful in the degradation of cellulose, and in generating specific anti-  
CC Aviii antibodies that are useful in purifying recombinant Aviii  
CC polypeptides from genetically engineered host cells, in detecting Aviii  
CC polypeptide expression, as well as a reagent tool for characterizing the  
CC molecular actions of the polypeptide. The Aviii polynucleotide is useful  
CC as a source of probes or primers in various diagnostic assays  
XX  
SQ Sequence 957 AA;

Query Match 98.5%; Score 462; DB 6; Length 957;  
Best Local Similarity 97.7%; Pred. No. 4.8e-44;  
Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VSGGVKQYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTYVWFTRDGSSSTLVYVNC 60  
DB 869 VSGGVKQYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTYVWFTRDGSSSTLVYVNC 928  
QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
DB 929 WAAIGCGNIRASFGSVNPATPTADTYLQ 956

RESULT 10  
ADD22921  
ID ADD22921 standard; protein; 957 AA.  
XX  
AC ADD22921;  
XX  
XX 15-JAN-2004 (first entry)  
XX  
DE Acidothermus cellulolyticus avicelase Aviii.  
XX  
KW enzyme; Aviii; cellulose reduction; agricultural biomass;  
KW municipal solid waste; glycoside hydrolase; avicelase.

OS Acidothermus cellulolyticus.  
XX  
XX Key Location/Qualifiers  
FT Misc-difference 957 /label= Unknown  
FT /note= "Encoded by N"  
XX  
XX US2003108988-A1.  
XX  
XX 12-JUN-2003.  
XX  
XX 18-OCT-2002; 2002US-00155400.  
XX  
XX 28-JUL-2001; 2001US-00917376.  
XX

PA (DING/) DING S.  
PA (ADNE/) ADNEY W S.  
PA (VINZ/) VINZANT T B.  
PA (HIMM/) HIMMEL M E.  
XX  
XX Ding S, Adney WS, Vinzant TB, Himmel ME;  
XX  
XX WPI; 2003-810853/76.  
XX N-PSDB; ADD22922.  
XX  
XX New isolated thermal tolerant avicelase polynucleotide useful for  
XX detection of a polynucleotide encoding Aviii and for reducing cellulose  
XX in a starting material, e.g. municipal solid waste.

PS Claim 16; SEQ ID NO 1; 29pp; English.  
XX  
XX The invention relates to an isolated polynucleotide molecule encoding a  
XX thermostable Aviii polypeptide. The polynucleotide is useful for  
XX detection of a polynucleotide encoding Aviii. The polynucleotide is  
XX useful for reducing cellulose in a starting material which involves



CC administering to the starting material, e.g. agricultural biomass or  
 CC municipal solid waste, a polypeptide molecule of the polynucleotide. The  
 CC method further comprises administering a second polypeptide molecule  
 CC chosen from the glycoside hydrolase family of proteins. The present  
 CC sequence represents the amino acid sequence of Acidothermus  
 CC cellulolyticus avicelase AvIII.  
 XX  
 SQ Sequence 957 AA;

Query Match 98.5%; Score 462; DB 7; Length 957;  
 Best Local Similarity 97.7%; Pred. No. 4.8e-44;  
 Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VSGGVKQYKNDSPAGDNQIKPGLQVNTGSSVDLSVTYVYFTRDGGSSTLVYNC 60  
 DB 869 VSGGVKQYKNDSPAGDNQIKPGLQVNTGSSVDLSVTYVYFTRDGGSSTLVYNC 928  
 QY 61 WAAACGNGIRASFGSVNPATPTADTYLQ 88  
 DB 929 WAAACGNGIRASFGSVNPATPTADTYLQ 956

RESULT 11  
 ABP71658  
 ID ABP71658 standard; protein; 153 AA.  
 AC ABP71658;  
 XX  
 DT 29-MAY-2003 (first entry)  
 XX  
 DE A. cellulolyticus Gux1 protein CBD\_III domain fragment.  
 XX  
 KW Gux1; thermal tolerant; exoglucanase; glycoside hydrolase; cellulose;  
 KW biofuel; detergent; pulp; paper processing; feed processing; textile;  
 KW cellulase; enzyme.  
 XX  
 OS Acidothermus cellulolyticus.  
 XX  
 PN WO2003012095-A1.  
 XX  
 PD 13-FEB-2003.

XX 28-JUL-2001; 2001WO-US023820.  
 XX 28-JUL-2001; 2001WO-US023820.  
 XX (MIDE ) MIDWEST RES INST.  
 XX Adney WS, Ding S, Vinzant TB, Himmel ME, Decker SR;  
 PI Lantz McCarter S;  
 PI WPI; 2003-300494/29.  
 DR  
 XX New thermal tolerant Gux1 peptide having specified amino acid sequence,  
 PT useful in the degradation of cellulose to biofuels.  
 PT  
 PS Claim 2; Page 7; 44pp; English.

XX The invention relates to a thermal tolerant Gux1 peptide from A.  
 CC cellulolyticus. The Gux1 exoglucanase is a member of the glycoside  
 CC hydrolase family and comprises a catalytic domain GH48, carbohydrate  
 CC binding domain type III, and a carbohydrate binding domain type II. The  
 CC polypeptide is useful in the degradation of cellulose into biofuel, or  
 CC for conversion of biomass to biofuel additives. It is used in detergents,  
 CC pulp and paper processing, food and feed processing, and in textile  
 CC processing. It can also be used alone or in combination with other  
 CC cellulase or glycoside hydrolases. The novel polypeptide generates  
 CC alternative cellulase enzymes capable of assisting in the commercial-  
 CC scale processing of cellulose to sugar for use in biofuel production. The  
 CC present sequence represents a A. cellulolyticus Gux1 cellulase CBD\_III  
 CC domain fragment  
 XX  
 XX Sequence 153 AA;

Query Match 93.0%; Score 436; DB 6; Length 153;  
 Best Local Similarity 96.4%; Pred. No. 4.9e-42;  
 Matches 81; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 5 VKQYKNDSPAGDNQIKPGLQVNTGSSVDLSVTYVYFTRDGGSSTLVYNCDAAM 64  
 DB 4 LKAQYKNDSPAGDNQIKPGLQVNTGSSVDLSVTYVYFTRDGGSSTLVYNCDAAM 63  
 QY 65 GCGNIRASFGSVNPATPTADTYLQ 88  
 DB 64 GCGNIRASFGSVNPATPTADTYLQ 87

RESULT 12  
 ABP71656  
 ID ABP71656 standard; protein; 1121 AA.  
 XX  
 AC ABP71656;  
 XX  
 DT 29-MAY-2003 (first entry)  
 XX  
 DE A. cellulolyticus Gux1 protein.  
 XX  
 KW Gux1; thermal tolerant; exoglucanase; glycoside hydrolase; cellulose;  
 KW biofuel; detergent; pulp; paper processing; feed processing; textile;  
 KW cellulase; enzyme.  
 XX  
 OS Acidothermus cellulolyticus.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..34 /note= "potential signal peptide"  
 FT Protein 35..1121 /note= "mature protein"  
 FT Misc-difference 228 /note= "encoded by CG"  
 FT  
 XX  
 PN WO2003012095-A1.  
 XX  
 PD 13-FEB-2003.

XX 28-JUL-2001; 2001WO-US023820.  
 XX 28-JUL-2001; 2001WO-US023820.  
 XX (MIDE ) MIDWEST RES INST.  
 XX Adney WS, Ding S, Vinzant TB, Himmel ME, Decker SR;  
 PI Lantz McCarter S;  
 PI WPI; 2003-300494/29.  
 DR N-PSDB; ABZ76162.  
 DR  
 XX New thermal tolerant Gux1 peptide having specified amino acid sequence,  
 PT useful in the degradation of cellulose to biofuels.  
 PT  
 PS Claim 3; Page 18-19; 44pp; English.

XX The invention relates to a thermal tolerant Gux1 peptide from A.  
 CC cellulolyticus. The Gux1 exoglucanase is a member of the glycoside  
 CC hydrolase family and comprises a catalytic domain GH48, carbohydrate  
 CC binding domain type III, and a carbohydrate binding domain type II. The  
 CC polypeptide is useful in the degradation of cellulose into biofuel, or  
 CC for conversion of biomass to biofuel additives. It is used in detergents,  
 CC pulp and paper processing, food and feed processing, and in textile  
 CC processing. It can also be used alone or in combination with other  
 CC cellulase or glycoside hydrolases. The novel polypeptide generates  
 CC alternative cellulase enzymes capable of assisting in the commercial-  
 CC scale processing of cellulose to sugar for use in biofuel production. The  
 CC present sequence represents a A. cellulolyticus Gux1 cellulase  
 XX  
 XX Sequence 1121 AA;

AA113494
ID AA113494 standard; protein; 616 AA.
XX AC AAC13494;
XX AC AAC13494;
XX DT 30-JUL-1999 (first entry)
XX DE Truncated cellulase Cel E3/B5.
XX KW Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;
KW KM Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;
KW KM cotton-containing fabric; stonewashing.
XX OS Unidentified.
XX PN EP921188-A2.
XX PD 09-JUN-1999.
XX PF 15-SEP-1998; 98EP-00810919.
XX PR 19-SEP-1997; 97US-00932571.
XX PA (CLRN ) CLARIANT FINANCE BVI LTD.
XX PI Anderson P, Bergquist PL, Daniels RM, Farrington GK, Gibbs MD;
PI PI Morgan H, Williams DP;
XX DR WPI; 1999-315403/27.
DR DR N-PSDB; AAX55660.
XX NT New truncated cellulase proteins, useful in detergents and for producing
PT PT 'stonewashed' denim.
XX PS Claim 7; Page 42-43; 65pp; English.
XX CC The invention relates to a recombinant cellulase active protein free of
CC CC proteinases of native thermophilic and alkaliphilic origin, comprising
CC CC the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3,
CC CC Cel 6 or Cel E3/B5, or a stability region from one of the defined full-
CC CC length sequences, or functional equivalents. Cel B5 extends from amino
CC CC acid A1011 to P1424 or K1425 or N1426, and Cel B4/5 extends from amino
CC CC acid K635 to N1426 in the sequence shown in AA113492; Cel E1 extends from
CC CC amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel E1/2/3
CC CC extends from Y39 to G812, Cel B6 extends from amino acid V1233 to K1751
CC CC and the stability region extends from amino acid S482 to G635 in the
CC CC sequence shown in AA113493; Cel E3/B5 is shown in AA113494. The new
CC CC enzymes are useful in laundry detergent compositions to prevent or remove
CC CC staining, backstaining or graying, for use on cellulosic materials
CC CC including cotton-containing fabrics. They are especially useful for
CC CC preventing redeposition of colorant during stonewashing, and for
CC CC processing of textiles where cellulose breakdown is required. The new
CC CC truncated enzymes show reduced redeposition of dye compared to using non
CC CC truncated cellulase compositions
XX SQ Sequence 616 AA;
Query Match 41.7%; Score 195.5; DB 2; Length 616;
Best Local Similarity 43.2%; Pred. No. 1.3e-13;
Matches 38; Conservative 15; Mismatches 34; Indels 1; Gaps
QY 1 VSGGVKVOVKNDSPAGDNQIKPGIQLVNTGSSVDLSTVTVRVWFTRDGGSSTLVYNCD 60
DB 1 MSGGVKVLKYNNNETASTGISIRPFKIVNGSSSVDDLGRVKIRVWTVDGDKPQSAY-CD 59
QY 61 WAAMGCCNIRASFGSVNPATPTADTYLQ 88
DB 60 WAQIGASNVTFEVLKSSGSGADIYLE 87
RESULT 15
AAE16325
ID AAE16325 standard; protein; 616 AA.



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 11:51:45 ; Search time 4.1265 Seconds  
(without alignments)  
2051.340 Million cell updates/sec

Title: US-09-917-376-5  
Perfect score: 469  
Sequence: 1 VSGGVKVOYKNNDSAPGNDQ.....IRASFGSVNPATPTADTYLQ 88  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues  
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	191	40.7	1711	2 T31337	1,4-beta-glucanase
2	190.5	40.6	1779	2 T31085	xylanase - Caldice
3	190	40.5	508	2 G63593	cellulase (EC 3.2.
4	190	40.5	508	2 A26874	cellulase (EC 3.2.
5	188	40.1	499	2 JN0111	cellulase (EC 3.2.
6	186.5	39.8	1331	2 A48954	mannan endo-1,4-be
7	186.5	39.8	1742	2 T17120	cellulase (EC 3.2.
8	186	39.7	499	2 A27198	cellulase (EC 3.2.
9	182.5	38.9	915	2 A43802	cellulase (EC 3.2.
10	182.5	38.9	1039	2 S02711	cellulase (EC 3.2.
11	180.5	38.5	145	2 A41897	cellulase homolog
12	180	38.4	486	2 I40548	bifunctional cellu
13	172	36.7	700	2 B41897	cellulase (EC 3.2.
14	165.5	35.3	879	2 A47704	endoglucanase I (E
15	165	35.2	1854	2 S36859	cipA protein - Clo
16	149.5	31.9	505	2 S39962	endoglucanase - Er
17	147.5	31.4	504	2 S54744	cellulase (EC 3.2.
18	141.5	30.2	986	2 S12021	thermoactive cellu
19	130.5	27.8	586	2 PC5006	scaffolding proteoi
20	126.5	27.0	1483	2 C97012	probably celluloso
21	126.5	27.0	1162	2 T30433	scaffolding protei
22	117	24.9	1230	2 T30433	cellulose 1,4-beta
23	114	24.3	1848	2 A44140	cellulose-binding
24	75.5	16.1	618	2 T08685	hypothetical prote
25	73.5	15.7	547	2 T25478	hypothetical prote
26	73.5	15.7	1428	2 AC2224	hypothetical prote
27	70	14.9	5188	2 B85547	probable RXR famil
28	70	14.9	5291	2 F90696	hypothetical prote
29	68.5	14.6	574	2 A69196	cell surface glyco

30	67	14.3	271	2 H72684	hypothetical prote
31	66.5	14.2	247	2 JC7201	thaumatin-like pro
32	66	14.1	535	2 T47790	hypothetical prote
33	65.5	14.0	751	2 T40462	ser-lys rich hypot
34	65.5	14.0	1435	2 T01075	polyprotein - hepa
35	65	13.9	523	2 C70717	probable purH prot
36	64	13.6	1222	2 G72614	probable reverse g
37	64	13.6	1571	2 S50669	hypothetical prote
38	64	13.6	1594	2 T30549	hemolysin - rabbit
39	63.5	13.5	398	2 H70393	hemolysin - Aquife
40	63.5	13.5	2403	2 A59386	sanko - human
41	62.5	13.3	269	2 H95976	endo-beta-1,3-1,4-
42	62.5	13.3	454	2 AB2512	replicative DNA he
43	62.5	13.3	802	1 B44390	protein-tyrosine-p
44	62.5	13.3	4936	2 AH2515	hypothetical prote
45	62	13.2	122	2 E37267	Ig heavy chain V r

ALIGNMENTS

RESULT 1

T31337  
1,4-beta-glucanase (EC 3.2.1.-) - Anaerocellum thermophilum (fragment)  
C:Species: Anaerocellum thermophilum  
C:Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 02-Sep-2000  
C:Accession: T31337  
R:Zverlov, V.; Mahr, S.; Riedel, K.; Bronnenmeier, K.  
Microbiology 144: 457-465, 1998  
A:Title: Properties and gene structure of a bifunctional cellulolytic enzyme (Cela) from omain.  
A:Reference number: Z21003; MUID:98154434; PMID:9493383  
A:Accession: T31337  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1711 <ZVS>  
A:Cross-references: EMBL:Z86105; NID:el071329; PID:e350354; PIDN:CAB06786.1  
C:Genetics:  
A:Gene: cela  
C:Keywords: glycosidase; hydrolase

Query Match	40.7%;	Score 191;	DB 2;	Length 1711;
Best Local Similarity	43.8%;	Pred. No. 1.2e-11;		
Matches	39;	Conservative 18;	Mismatches 30;	Indels 2; Gaps 2;
QY	1	VSGG-VKVOYKNNDSAPGNDQIKPGIQLVNTGSSVDLSTVTYRYWETRGSGSTLYVNC	59	
DB	683	VAGGQIKVLYANKETNSTTIRPWLKVNVTGSSIDLRSVTIRYWTVDGKAQSAIS-	741	
QY	60	DWAAMCGNIRASFQSVNPATPTADTYLQ	88	
DB	742	DWAQIGASNVTFKVLSSVSGADYILE	770	

RESULT 2

T31085  
xylanase - Caldicellulosiruptor sp.  
C:Species: Caldicellulosiruptor sp.  
C:Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 02-Sep-2000  
C:Accession: T31085  
R:Morris, D.D.; Gibbs, M.D.; Ford, M.; Thomas, J.; Bergquist, P.L.  
submitted to the EMBL Data Library, December 1997  
A:Description: Family 10 and 11 xylanase genes from Caldicellulosiruptor sp. Rt69B.1.  
A:Reference number: Z20972  
A:Accession: T31085  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1779 <MOR>  
A:Cross-references: EMBL:AF036924; NID:g2760905; PID:g2760909; PIDN:AAB95326.1  
C:Genetics:  
A:Note: xyc

Query Match 40.6%; Score 190.5; DB 2; Length 1779;

[illegible]

```

QY      2  SGGVKVQYKXNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRVYFWTRDGSSTLVYNCW 61
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db     364  SGQIKVLYANKETNTTIRPWLKVVNGSSSIDLSRVTIRYWTVDGERQAIS-DW 422

QY      62  AAMCGCNIRASFGSVNPATPTADTYLQ 88
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db     423  AQIGASNVTFKFKVGLSSVSGADYYLE 449

RESULT 7
T17120
cellulase (EC 3.2.1.1) precursor, thermoactive - Caldocellum saccharolyticum
C/Species: Caldocellum saccharolyticum
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Nov-2000
C/Accession: T17120; A43745
R;Te'o V.S.; Saul, D.J.; Bergquist, P.L.
Appl. Microbiol. Biotechnol. 43, 291-296, 1995
A;Title: Cella, another gene coding for a multidomain cellulase from the extreme
A;Reference number: Z19698; MUID:95336703; PMID:7612247
A;Accession: T17120
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1742 <TEO>
A;Cross-references: EMBL:L32742; NID:G537499; PID:G537500; PIDN:AAA91086.1
R;Luethi, E.; Bhana Jagmat, N.; Grayling, R.A.; Love, D.R.; Bergquist, P.L.
Appl. Environ. Microbiol. 57, 694-700, 1991
A;Title: Cloning, sequence analysis, and expression in Escherichia coli of a gene
A;Reference number: A43745; MUID:91247819; PMID:2039230
A;Accession: A43745
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1516-1544, 'A', 1546-1742 <LUB>
A;Cross-references: EMBL:M36063; NID:G144292; PIDN:AAA72860.1; PID:G144293
C;Genetics:
A;Gene: celsA
A;Keywords: glycosidase; hydrolase

Query Match          39.8%; Score 186.5; DB 2; Length 1742;
Best Local Similarity 42.5%; Pred. No. 3.7e-11;
Matches 37; Conservative 18; Mismatches 31; Indels - 1; Gaps 1;

QY      2  SGGVKVQYKXNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRVYFWTRDGSSTLVYNCW 61
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db     704  SGQIKVLYANKETNTTIRPWLKVVNGSSSIDLSRVTIRYWTVDGERQAIS-DW 762

QY      62  AAMCGCNIRASFGSVNPATPTADTYLQ 88
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db     763  AQIGASNVTFKFKVGLSSVSGADYYLE 789

RESULT 8
A27198
cellulase (EC 3.2.1.4) precursor - Bacillus subtilis (strain IFO3034)
N/Alternate names: endo-1,4-beta-glucanase
C/Species: Bacillus subtilis
C/Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 15-Oct-1999
C/Accession: A27198
R;Nakamura, A.; Uozumi, T.; Beppu, T.
Eur. J. Biochem. 164, 317-320, 1987
A;Title: Nucleotide sequence of a cellulase gene of Bacillus subtilis.
A;Reference number: A27198; MUID:87190397; PMID:3106035
A;Accession: A27198
A;Molecule type: DNA
A;Residues: 1-499 <NAB>
A;Cross-references: GB:M29332; NID:G142670; PIDN:AAA22307.1; PID:G142671
A;Experimental source: strain IFO3034
C;Function:
A;Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans
A;Pathway: cellulose degradation
C;Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degr
F;1-36/Domain: signal sequence #status predicted <SIG>

```

```

F;29-1039/Product: cellulase #status predicted <MAT>
F;72-373/Domain: Streptomyces endo-1,4-beta-xylosidase A homology <SKY>

Query Match          38.9%; Score 182.5; DB 2; Length 1039;
Best Local Similarity 41.9%; Pred. No. 5.5e-11;
Matches 36; Conservative 18; Mismatches 31; Indels 1; Gaps 1;

Qy 3 GGVKVOYKNNDSPGDNQIKPGQLQVNTGSSVDLSTVTYRYWFTRDGSGSTLYVNCDDWA 62
Db 420 GQIKVLVYANKETSTNTTIRPWLKVNVSGSSIDLSTRTYRYWFTVDGERAQSAVS-DWA 478

Qy 63 AMCGNIRASFGSVNPATPTADTYLQ 88
Db 479 QIGASNVTFKFKVLSSVSGADYYLE 504

RESULT 11
A41897
cellulase homolog - Bacillus lautus (fragment)
C/Species: Bacillus lautus
C/Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 15-Oct-1999
C/Accession: A41897; S27498
R/Hansen, C.K.; Diderichsen, B.; Jorgensen, P.L.
J. Bacteriol. 174, 3522-3531, 1992
A/Title: celiA from Bacillus lautus PL236 encodes a novel cellulose-binding endo
A/Reference number: A41897; MUID:92276330; PMID:1592807
A/Accession: A41897
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-145 <HAN>
A/Cross-references: EMBL:M76588; NID:g142661; PIDN:AAA22302.1; PID:g142662
A/Experimental source: PL236
A/Note: sequence extracted from NCBI backbone (NCBIP:104604)

Query Match          38.5%; Score 180.5; DB 2; Length 145;
Best Local Similarity 43.9%; Pred. No. 1.1e-11;
Matches 36; Conservative 18; Mismatches 25; Indels 3; Gaps 2;

Qy 7 VOYKNNDSPGDNQIKPGQLQVNTGSSVDLSTVTYRYWFTRDGSGSTLYVNCDDWAAMGC 66
Db 1 LQPRADTNAADNQIKPSFNKNGTISAVDLSTLKTRYFTKDGSAVNGW-IDWAQLGG 59

Qy 67 GNIRASFGSVNPATPTADTYLQ 88
Db 60 SNIOISFG-NHTGTNSDTWE 79

RESULT 12
I40548
bifunctional cellulase precursor - Bacillus sp.
C/Species: Bacillus sp.
C/Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 15-Oct-1999
C/Accession: I40548
R/Han, S.J.; Yoo, Y.J.; Kang, H.S.
J. Biol. Chem. 270, 26012-26019, 1995
A/Title: Characterization of a bifunctional cellulase and its structural gene:
A/Reference number: I40548; MUID:96029707; PMID:7592793
A/Accession: I40548
A/Status: preliminary; translated from GB/EMBL/DDBJ
A/Molecule type: DNA
A/Residues: 1-486 <RES>
A/Cross-references: EMBL:U27084; NID:g857575; PIDN:AAC43478.1; PID:g857576
C/Genetics:
A/Gene: cel

Query Match          38.4%; Score 180; DB 2; Length 486;
Best Local Similarity 38.8%; Pred. No. 4.5e-11;
Matches 33; Conservative 18; Mismatches 32; Indels 2; Gaps 1;

Qy 4 GVKVOYKNNDSPGDNQIKPGQLQVNTGSSVDLSTVTYRYWFTRDGSGSTLYVNCDDWA 63
Db 353 GISVQYRAGDGSMNSNQIRPOLQIKNGNTVDLKDVTARYWYNAKNKGQNL--DCDYEQ 410

```

QY 64 MCGNIRASFGSVNPAFTADTYLQ 88  
DB 411 LGCGNVSTVTLHKPKQGADTYLE 435

RESULT 13  
B41897  
cellulase (EC 3.2.1.4) - Bacillus lautus  
N:Alternate names: endo-1,4-beta-glucanase  
C:Species: Bacillus lautus  
C:Date: 10-Mar-1994 #sequence\_revision 10-Mar-1994 #text\_change 15-Oct-1999  
R:Accession: B41897; S27499  
J: Hansen, C.K.; Diderichsen, B.; Jorgensen, P.L.  
J. Bacteriol. 174, 3522-3531, 1992  
A:Title: Cella from Bacillus lautus PL236 encodes a novel cellulose-binding endo-beta-1,4  
A:Reference number: A41897; MUID:92276330; PMID:1592807  
A:Accession: B41897  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-700 <HAN>  
A:Cross-references: EMBL:M76588; NID:g142661; PIDN:AAA22303.1; PID:g142663  
A:Experimental source: PL236  
A:Note: sequence extracted from NCBI backbone (NCBIP:104605)  
C:Function:  
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce  
A:Pathway: cellulose degradation  
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 36.7%; Score 172; DB 2; Length 700;  
Best Local Similarity 44.9%; Pred. No. 4.5e-10;  
Matches 40; Conservative 18; Mismatches 27; Indels 4; Gaps 3;

QY 1 VSGGVKVOYKND-SAPGDNOIKPGLQLVNTGSSVDLSTVTVRYWFTDGGSSTLVYNC 59  
DB 549 VNSDLVVOYKDGDRNATDNOIKPHFNQNGTSPVLSLTLRYFTKD-SSAAMNGWI 607

QY 60 DWAMCGCNIRASFGSVNPAFTADTYLQ 88  
DB 608 DWAKLGGNIOISFGNHGA--DSDTYAE 634

RESULT 14  
A47704  
endoglucanase I (EC 3.2.1.-) Cell - Clostridium thermocellum  
C:Species: Clostridium thermocellum  
C:Date: 19-Dec-1993 #sequence\_revision 18-Nov-1994 #text\_change 15-Oct-1999  
R:Accession: A47704  
J: Hazlewood, G.P.; Davidson, K.; Laurie, J.I.; Huskisson, N.S.; Gilbert, H.J.  
J. Gen. Microbiol. 139, 307-316, 1993  
A:Title: Gene sequence and properties of CellI, a family E endoglucanase from Clostridium  
A:Reference number: A47704; MUID:93171873; PMID:8436949  
A:Accession: A47704  
A:Status: preliminary  
A:Molecule type: DNA; protein  
A:Residues: 1-879 <HAZ>  
A:Cross-references: GB:L04735; NID:g144807; PIDN:AAA20892.1; PID:g144808  
A:Note: sequence extracted from NCBI backbone (NCBIN:125637, NCBIP:125638)  
C:Keywords: glycosidase; hydrolase

Query Match 35.3%; Score 165.5; DB 2; Length 879;  
Best Local Similarity 33.0%; Pred. No. 2.7e-09;  
Matches 29; Conservative 26; Mismatches 32; Indels 1; Gaps 1;

QY 1 VSGGVKVOYKND-SAPGDNOIKPGLQLVNTGSSVDLSTVTVRYWFTDGGSSTLVYNC 60  
DB 736 IKGEVVLQYANGAGATSNINPRKIIINGTKAINLSDVIRYYTKEGGASQNFW-CD 794

QY 61 WAMCGCNIRASFGSVNPAFTADTYLQ 88  
DB 795 WSSAGNSNVGTGNFFNLSPKEGADTCL 822

RESULT 15

S36859  
cipA protein - Clostridium thermocellum  
N:Alternate names: probable cellulosome protein large chain SL  
C:Species: Clostridium thermocellum  
C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 01-Dec-2000  
R:Accession: S36859; S33527; S25767; S28659; T18261  
J: Gerengross, U.T.; Demain, A.L.  
submitted to the EMBL Data Library, January 1993  
A:Reference number: S36859  
A:Accession: S36859  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1854 <GER>  
A:Cross-references: EMBL:L08665  
R:Gerengross, U.T.; Romaniec, M.P.M.; Kobayashi, T.; Huskisson, N.S.; Demain, A.L.  
Mol. Microbiol. 8, 325-334, 1993  
A:Title: Sequencing of a Clostridium thermocellum gene (cipA) encoding the cellulosomal  
A:Reference number: S33527; MUID:93302508; PMID:8316083  
A:Accession: S33527  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1615,1617-1854 <GE2>  
A:Cross-references: EMBL:L08665  
R:Fujino, T.; Beguin, P.; Aubert, J.P.  
FEMS Microbiol. Lett. 94, 165-170, 1992  
A:Title: Cloning of a Clostridium thermocellum DNA fragment encoding polypeptides that  
A:Reference number: S25767  
A:Accession: S25767  
A:Molecule type: DNA  
A:Residues: 1216-1232, 'Y', 1234-1241, 'T', 1243-1319, 'R', 1321-1615, 1617-1854 <FUJ>  
R: Beguin, P.  
submitted to the EMBL Data Library, August 1992  
A:Reference number: S28659  
A:Accession: S28659  
A:Molecule type: DNA  
A:Residues: 1216-1232, 'Y', 1234-1241, 'T', 1243-1319, 'R', 1321-1615, 1617-1768, 'R', 1770-1854  
R: Fujino, T.; Beguin, P.; Aubert, J.P.  
J. Bacteriol. 175, 1891-1899, 1993  
A:Title: Organization of a Clostridium thermocellum gene cluster encoding the cellulosome  
e.  
A:Reference number: Z18847; MUID:93209931; PMID:8458832  
A:Accession: T18261  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1821-1854 <FU2>  
A:Cross-references: EMBL:X67506; NID:g296879; PID:g296880; PIDN:CAA47840.1  
C:Genetics:  
A:Gene: cipA

Query Match 35.2%; Score 165; DB 2; Length 1854;  
Best Local Similarity 36.1%; Pred. No. 6.9e-09;  
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKVOYKND-SAPGDNOIKPGLQLVNTGSSVDLSTVTVRYWFTDGGSSTLVYNC 60  
DB 365 VSGNLKVEYFNPSPTTNSINPQFKVTNTGSSAIDLKLTLYRYTVDGQKQOTFW-CD 423

QY 61 WAMCGCNIRASFGSVNPAFTADTYLQ 88  
DB 424 HAAITGNSNGSYNGITSNVKGTFVKMSSSTNNADTYLE 460

Search completed: May 11, 2004, 12:10:26  
Job time : 4.1265 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 11:41:35 ; Search time 2.39913 Seconds  
(without alignments)  
1909.933 Million cell updates/sec

Title: US-09-917-376-5

Perfect score: 469

Sequence: 1 VSGGVKQYKNDSPAGDQ.....IRASFGSVNPATPTADTYLQ 88

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	190	40.5	499	1 GUN1_BACSU	P07983 bacillus su
2	190	40.5	499	1 GUN2_BACSU	P10475 bacillus su
3	188	40.1	499	1 GUN3_BACSU	P23549 bacillus su
4	186.5	39.8	1331	1 MANB_CALSA	P22533 caldocellum
5	186.5	39.8	1742	1 GUN4_CALSA	P22534 caldocellum
6	182.5	38.9	1039	1 GUNB_CALSA	P10474 c endogluc
7	180.5	38.5	145	1 YCEA_PAEIA	P29718 paenibacill
8	172	36.7	700	1 GUN4_PAEIA	P02934 clostridium
9	165.5	35.3	879	1 GUN1_CLOTM	P29719 paenibacill
10	165	35.2	772	1 CIPB_CLOTM	Q01866 clostridium
11	165	35.2	1853	1 CIPB_CLOTM	Q06851 clostridium
12	149.5	31.9	444	1 GUNV_ERWCA	Q59394 erwinia car
13	149.5	31.9	505	1 GUNV_ERWCA	Q47096 erwinia car
14	147.5	31.4	504	1 GUNV_ERWCA	Q59395 erwinia car
15	147.5	31.4	914	1 GUX2_CLOS	P50900 clostridium
16	141.5	30.2	986	1 GUNZ_CLOS	P23659 clostridium
17	114	24.3	1848	1 CBPA_CLOCL	P38058 clostridium
18	67.5	14.4	118	1 HV39_MOUSE	P01809 mus musculus
19	67.5	14.4	215	1 FLA1_METVA	P95316 methanococc
20	66.5	14.2	246	1 TP1A_MALDO	Q9fsg7 malus domes
21	65	13.9	523	1 PUR9_MYCTU	P71553 h bifunctio
22	64.5	13.8	556	1 ILVD_THETN	Q8rdj9 thermoanear
23	64	13.6	1571	1 ATCS_YEAST	P32660 saccharomyc
24	63.5	13.5	247	1 FLA1_THEVO	P57719 thermoplasm
25	63.5	13.5	465	1 TY3H_SCHMA	O17446 schistosoma
26	63	13.4	953	1 O106_HUMAN	Q9upv9 homo sapien
27	63	13.4	1581	1 PRB8_HUMAN	Q15648 h peroxisom
28	62.5	13.3	269	1 EXOK_RHIME	P33693 rhizobium m
29	62.5	13.3	802	1 PTN8_MOUSE	P29352 mus musculus
30	62	13.2	950	1 STL2_MOUSE	Q99n50 mus musculus
31	62	13.2	1210	1 ICEN_PSEPL	P09815 pseudomonas
32	61.5	13.1	252	1 HIS6_RHOSH	P50937 rhodobacter
33	61.5	13.1	316	1 MUCB_PSEAE	P38108 pseudomonas

34	61.5	13.1	328	1 HAIQ_MOUSE	P14428 mus musculus
35	61.5	13.1	368	1 HAIW_MOUSE	P03991 mus musculus
36	61.5	13.1	489	1 GATA_CAUCR	Q9a510 caulobacter
37	61.5	13.1	678	1 YIHQ_ECOLI	P32138 escherichia
38	61	13.0	120	1 HV50_MOUSE	P06329 mus musculus
39	61	13.0	338	1 RTCA_ECOL6	O8fca8 escherichia
40	61	13.0	338	1 RTCA_ECOLI	P46849 escherichia
41	61	13.0	342	1 RTCA_ECO57	P58127 escherichia
42	60.5	12.9	1157	1 C9CA_BACTO	Q45733 bacillus th
43	60	12.8	312	1 LECF_ALEAU	P18891 aleuria aur
44	60	12.8	1226	1 PAT2_CAEEL	P34446 caenorhabdi
45	59.5	12.7	261	1 COTN_BACSU	P54507 bacillus su

#### ALIGNMENTS

RESULT 1  
GUN1\_BACSU  
ID GUN1\_BACSU STANDARD; PRT; 499 AA.  
AC P07983;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-NOV-1991 (Rel. 20, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)  
GN (Cellulase).  
GN BGLC OR GLD.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DLG;  
RX MEDLINE=87194581; PubMed=3106328;  
RA Robson L.M., Chambliss G.H.;  
RL "Endo-beta-1,4-glucanase gene of Bacillus subtilis DLG.";  
RL J. Bacteriol. 169:2017-2025(1987).  
CC -|- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic linkages in cellulose, lichenin and cereal beta-D-glucans.  
CC -|- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl hydrolases).

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
-----

EMBL; M16185; AAA22496.1; ALT\_INIT.  
PIR; A26874; A26874.  
HSSP; O85465; 1A3H.  
InterPro; IPR001956; CBD 3.  
InterPro; IPR008965; Cellul\_bind.  
InterPro; IPR001547; Glyco\_hydro\_5.  
Pfam; PF00942; CBM\_3; 1.  
Pfam; PF00150; cellulase; 1.  
ProDom; PD001947; CBD 3; 1.  
PROSITE; PS00559; GLYCOSYL HYDROL F5; 1.  
Cellulose degradation; Hydrolase; Glycosidase; Signal.  
FT SIGNAL 1 29  
FT CHAIN 30 499 ENDOGLUCANASE.  
FT ACT SITE 169 169 PROTON DONOR (BY SIMILARITY).  
FT ACT SITE 257 257 NUCLEOPHILE (BY SIMILARITY).  
FT DOMAIN 350 499 CELLULOSE-BINDING (BY SIMILARITY).  
SQ SEQUENCE 499 AA; 55187 MW; 339D04BE95A63EE1 CRC64;

Query Match 40.5%; Score 190; DB 1; Length 499;  
Best Local Similarity 43.0%; Pred. No. 1.3e-13;  
Matches 37; Conservative 17; Mismatches 28; Indels 4; Gaps 2;

OY 4 GVKVQKNDSPAGDQIKPGIQLVNTGSSVDLSTVTYVWF-TRDGGSTLYVNCDA 62

```

Db 353 GVSQYKAGDGVNSQIRPOLHKNNGNATVDLKDVTARYWYVKNKNGQN---FDCDYA 409
QY 63 AMGCGNIRASFGSVNPAATPTADTYLQ 88
Db 410 QMCGCNLTHKFTVTLHKPKQGADTYLE 435

RESULT 2
GUN2_BACSU
ID GUN2_BACSU STANDARD; PRT; 499 AA.
AC P10475;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Carboxymethyl-cellulase) (CMCANE) (Cellulase).
GN BGIC OR GLD OR EGLS OR ESU18130.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PA115;
RX MEDLINE=87066783; PubMed=3024130;
RA Mackay R.M., Lo A., Willick G., Zuker M., Baird S., Dove M.,
RA Moranelli F., Seligy V.;
RT "Structure of a Bacillus subtilis endo-beta-1,4-glucanase gene.";
RL Nucleic Acids Res. 14:9159-9170(1986).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CK-2;
RX MEDLINE=95225656; PubMed=7710280;
RA Lindahl V., Aa K., Tronsmo A.;
RT "Nucleotide sequence of an endo-beta-1,4-glucanase gene from Bacillus
RT subtilis CK-2.";
RL Antonie Van Leeuwenhoek 66:327-332(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=97124194; PubMed=8969507;
RA Rose M., Entian K.D.;
RT "New genes in the 170 degrees region of the Bacillus subtilis genome
RT encode DNA gyrase subunits, a thioredoxin, a xylanase and an amino
RT acid transporter.";
RL Microbiology 142:3097-3101(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bartoro M.G., Bessieres P., Bolestin A., Borchert S.,
RA Bourissier R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerston I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dunsthof A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrati E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghm S.Y., Glaser P., Goffeau A., Gough E.J., Grandi G.,
RA Guseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Serror S.J., Serror P., Shin B.S., Solido B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,

```

```

RA Takeuchi M., Tamakoshi A., Tanaka T., Terpetra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassaretti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256(1997).
RN [5]
RP SEQUENCE OF 30-45.
RC STRAIN=CK-2;
RX MEDLINE=95225655; PubMed=7710279;
RA Aa K., Flengsrud R., Lindahl V., Tronsmo A.;
RT "Characterization of production and enzyme properties of an
RT endo-beta-1,4-glucanase from Bacillus subtilis CK-2 isolated from
RT compost soil.";
RL Antonie Van Leeuwenhoek 66:319-326(1994).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl
CC hydrolases).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z29076; CAA82317.1; -
DR EMBL; X04689; CAA28392.1; -
DR EMBL; X67044; CAA47429.1; -
DR EMBL; Z73234; CAA97610.1; ALT_INIT.
DR EMBL; Z99113; CAB13696.1; ALT_INIT.
DR PIR; G69593; G69593.
DR HSP; O85465; 1A3H.
DR Subtilist; BG10437; bg1C.
DR InterPro; IPR001956; bglC.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR001547; Glyco_Hydro_5.
DR Pfam; PF00942; CBM 3; 1.
DR Pfam; PF00150; cellulase; 1.
DR ProDom; PD001947; CBD 3; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal;
KW Complete proteome.
FT SIGNAL 1 29
FT CHAIN 30 499 ENDOGLUCANASE.
FT ACT_SITE 169 169 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 257 257 NUCLEOPHILE (BY SIMILARITY).
FT DOMAIN 350 499 CELLULOSE-BINDING (BY SIMILARITY).
FT CONFLICT 283 283 S -> N (IN REF. 2).
SQ SEQUENCE 499 AA; 55287 MW; 8F735FF711B3EAE2 CRC64;

Query Match 40.5%; Score 190; DB 1; Length 499;
Best Local Similarity 40.7%; Pred. No. 13e-13;
Matches 35; Conservative 21; Mismatches 26; Indels 4; Gaps 2;

QY 4 GVKVQYKNDASPGDNOIRPGLQVNTGSSVSLSTVTVYWF-TRDGGSTLYNCDWA 62
Db 353 GISVQYRAGDGSNNSQIRPOLQIKNGNTVLDKDVTRYWYVKNKNGQN---FDCDYA 409
QY 63 AMGCGNIRASFGSVNPAATPTADTYLQ 88
Db 410 QMCGCNLTHKFTVTLHKPKQGADTYLE 435

RESULT 3
GUN3_BACSU
ID GUN3_BACSU STANDARD; PRT; 499 AA.
AC P23549;
DT 01-NOV-1991 (Rel. 20, Created)

```

DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DE 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)  
 DE (Carboxymethyl-cellulase) (CMCase) (Cellulase).  
 GN BGLC.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BSB616;  
 RX MEDLINE=91299280; PubMed=1368694;  
 RA Park S.H., Kim H.K., Pack M.Y.;  
 RT "Characterization and structure of the cellulase gene of Bacillus subtilis BSB616";  
 RL Agric. Biol. Chem. 55:441-448(1991).  
 CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic linkages in cellulose, lichenin and cereal beta-D-glucans.  
 CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl hydrolases).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; D01057; BAA0859.1; -;  
 CC PIR; JN0111; JN0111.  
 CC HSSP; O85465; 1A3H.  
 CC InterPro; IPR001956; CBD 3.  
 CC InterPro; IPR008965; Cellul\_bind.  
 CC InterPro; IPR001547; Glyco\_hydro\_5.  
 CC Pfam; PF00942; CBM\_3; 1.  
 CC Pfam; PF00150; cellulase; 1.  
 CC ProDom; PD001947; CBD 3; 1.  
 CC ProSITE; PS00659; GLYCOSYL\_HYDROL\_F5; 1.  
 CC Cellulose degradation; Hydrolase; Glycosidase; Signal.  
 KW SIGNAL 1 29  
 FT CHAIN 30 499 ENDOGLUCANASE.  
 FT ACT\_SITE 169 169 PROTON DONOR (BY SIMILARITY).  
 FT ACT\_SITE 257 257 NUCLEOPHILE (BY SIMILARITY).  
 FT DOMAIN 350 499 CELLULOSE-BINDING (BY SIMILARITY).  
 SQ SEQUENCE 499 AA; 55169 MW; 2E21E3DBBBACA04 CRC64;  
 Query Match 40.1%; Score 188; DB 1; Length 499;  
 Best Local Similarity 40.0%; Pred. No. 2.2e-13;  
 Matches 34; Conservative 18; Mismatches 31; Indels 2; Gaps 1;  
 QY 4 GYKVOYKNDAPGDNQIKPGLQVLVNTGSSVDLSLTVTVYVFTRDGSGSLTVYVNCWAA 63  
 DB 353 GSVQYRAGDSMNSNRIPQLQIKNGNTVTLKDVTVYVYNAKNGQNV--DCDYAQ 410  
 QY 64 MCGNIRAFSGVNPATPTADTYLQ 88  
 DB 411 LCGGNVTKFVTLHKPKQGADTYLE 435  
 RESULT 4  
 MANB\_CALSA  
 ID MANB\_CALSA STANDARD; PRT; 1331 AA.  
 AC P22533;  
 DT 01-AUG-1991 (Rel. 19, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Beta-mannanase/endoglucanase A precursor [Includes: Mannan endo-1,4-beta-mannosidase A (EC 3.2.1.78) (Beta-mannanase) (Endo-1,4-mannanase); Endo-1,4-beta-glucanase (EC 3.2.1.4) (Cellulase)].  
 GN MANA.  
 OS Caldocellum saccharolyticum (Caldicellulosiruptor saccharolyticus).  
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;

OC Caldicellulosiruptor.  
 OX NCBI\_TaxID=44001;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93119139; PubMed=1476429;  
 RA Gibbs M.D., Saul D.J., Luthi E., Bergquist P.L.;  
 RT "The beta-mannanase from 'Caldocellum saccharolyticum' is part of a multidomain enzyme";  
 RL Appl. Environ. Microbiol. 58:3864-3867(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91247819; PubMed=2039230;  
 RA Luehi E., Jasmat N.B., Grayling R.A., Love D.R., Bergquist P.L.;  
 RT "Cloning, sequence analysis, and expression in Escherichia coli of a gene coding for a beta-mannanase from the extremely thermophilic bacterium 'Caldocellum saccharolyticum'";  
 RL Appl. Environ. Microbiol. 57:694-700(1991).  
 CC -!- FUNCTION: DEGRADATION OF HEMICELLULOSES, THE SECOND MOST ABUNDANT POLYSACCHARIDES IN NATURE. CONTAINS TWO CATALYTIC DOMAINS WITH MANNANASE AND ENDOGLUCANASE ACTIVITIES.  
 CC -!- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-beta-D-mannosidic linkages in mannans, galactomannans, glucomannans, and galactoglucomannans.  
 CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic linkages in cellulose, lichenin and cereal beta-D-glucans.  
 CC -!- MISCELLANEOUS: This enzyme is most active at pH 6 and 80 degrees Celsius.  
 CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL HYDROLASES).  
 CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY J (FAMILY 44 OF GLYCOSYL HYDROLASES).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; L01257; AAA71887.1; -;  
 CC EMBL; M36063; AAA72861.1; -;  
 CC PIR; A48954; A48954.  
 CC HSSP; O06851; INBC.  
 CC InterPro; IPR001956; CBD 3.  
 CC InterPro; IPR008965; Cellul\_bind.  
 CC InterPro; IPR001547; Glyco\_hydro\_5.  
 CC Pfam; PF00942; CBM 3; 2.  
 CC Pfam; PF00150; cellulase; 1.  
 CC ProDom; PD001947; CBD 3; 2.  
 CC ProSITE; PS00659; GLYCOSYL\_HYDROL\_F5; 1.  
 CC Hydrolase; Glycosidase; Cellulose degradation; Signal;  
 KW Multifunctional enzyme.  
 FT SIGNAL 1 41 POTENTIAL.  
 FT CHAIN 42 1331 BETA-MANNANASE/ENDOGLUCANASE A.  
 FT DOMAIN 42 325 CATALYTIC (MANNANASE ACTIVITY).  
 FT DOMAIN 326 361 PRO/SER/THR-RICH (PT BOX).  
 FT DOMAIN 362 518 SUBSTRATE-BINDING (POTENTIAL).  
 FT DOMAIN 519 564 PRO/SER/THR-RICH (PT BOX).  
 FT DOMAIN 565 720 SUBSTRATE-BINDING (POTENTIAL).  
 FT DOMAIN 721 780 PRO/SER/THR-RICH (PT BOX).  
 FT DOMAIN 781 1331 CATALYTIC (ENDOGLUCANASE ACTIVITY).  
 FT ACT\_SITE 162 162 PROTON DONOR (BY SIMILARITY).  
 FT ACT\_SITE 257 257 NUCLEOPHILE (BY SIMILARITY).  
 FT CONFLICT 338 338 T -> P (IN REF. 2).  
 FT CONFLICT 340 346 TPTPT -> RQHQHQ (IN REF. 2).  
 SQ SEQUENCE 1331 AA; 146892 MW; FFBCA51BB8DF0E0 CRC64;  
 Query Match 39.8%; Score 186.5; DB 1; Length 1331;  
 Best Local Similarity 42.5%; Pred. No. 1e-12;  
 Matches 37; Conservative 18; Mismatches 31; Indels 1; Gaps 1;  
 QY 2 SGVKVQYKNDAPGDNQIKPGLQVLVNTGSSVDLSLTVTVYVFTRDGSGSLTVYVNCW 61

```

Db 364 SQGIKVLANKETNTTIRPWLKVNSSGSSIDLSTRTIRYWTVDGERAQSAIS-DW 422
QY 62 AAMGCGNIRASFGSVNPTPTADTYLQ 88
Db 423 AQIGASNVTFKVLSSSVSGADYILE 449

RESULT 5
GUNA_CALSA
ID GUNA_CALSA STANDARD; PRT; 1742 AA.
AC P22534;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase A)
DE (Cellulase A).
GN CELA.
OS Caldocellum saccharolyticum (Caldicellulosiruptor saccharolyticus).
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=44001;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95336703; PubMed=7612247;
RA Te'O V.S., Saul D.J., Bergquist P.L.;
RT "celA, another gene coding for a multidomain cellulase from the
RT extreme thermophile Caldocellum saccharolyticum.";
RL Appl. Microbiol. Biotechnol. 43:291-296(1995).
RN [2]
RP SEQUENCE OF 1516-1742 FROM N.A.
RX MEDLINE=91247819; PubMed=2039230;
RA Leuthi E., Jasmat N.B., Grayling R.A., Love D.R., Bergquist P.L.;
RT "Cloning, sequence analysis, and expression in Escherichia coli of a
RT gene coding for a beta-mannanase from the extremely thermophilic
RT bacterium 'Caldocellum saccharolyticum'.";
RL Appl. Environ. Microbiol. 57:694-700(1991).
CC -!- FUNCTION: THE N-TERMINAL DOMAIN OF CELA ENCODES FOR AN
CC ENDOGLUCANASE ACTIVITY ON CARBOXYMETHYLCELLULOSE. THE C-TERMINAL
CC DOMAIN PROBABLY ACT SYNERGISTICALLY TO HYDROLYZE CRYSTALLINE
CC CELLULOSE.
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- PTM: The linker region (also termed "hinge") may be a potential
CC site for proteolysis.
CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
CC E (FAMILY 9 OF GLYCOSYL HYDROLASES).
CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
CC L (FAMILY 48 OF GLYCOSYL HYDROLASES).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L32742; AAA91086.1; -
CC EMBL; M36063; AAA72860.1; -
CC EMBL; L01257; -; NOT_ANNOTATED_CDS.
CC PIR; T17120; T17120.
CC HSP; P26221; 1TP4.
CC InterPro; IPR001956; CBD_3.
CC InterPro; IPR008965; Cellul_bind.
CC InterPro; IPR000556; Glyco_hydro_48.
CC InterPro; IPR001701; Glyco_hydro_9.
CC InterPro; IPR008928; Glyco_trans_6hp.
CC Pfam; PF00942; CBM_3; 3.
CC Pfam; PF02011; Glyco_hydro_48; 1.
CC Pfam; PF00759; Glyco_hydro_9; 1.
CC PRINTS; PR00844; GLHYDRLASE48.
CC ProDom; PD001947; CBD_3; 2.

```

---

```

DR ProDom; PD011903; Glyco_hydro 48; 1.
DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal; Repeat.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 1742 ENDOGLUCANASE A.
FT DOMAIN 24 642 CATALYTIC 1.
FT DOMAIN 643 700 LINKER ("HINGE") (PRO-THR BOX).
FT DOMAIN 701 857 CELLULOSE-BINDING (BY SIMILARITY).
FT DOMAIN 858 903 LINKER ("HINGE") (PRO-THR BOX).
FT DOMAIN 904 1060 CELLULOSE-BINDING (BY SIMILARITY).
FT DOMAIN 1061 1112 LINKER ("HINGE") (PRO-THR BOX).
FT DOMAIN 1113 1742 CATALYTIC 2.
FT ACT_SITE 396 396 BY SIMILARITY.
FT ACT_SITE 434 434 BY SIMILARITY.
FT ACT_SITE 443 443 BY SIMILARITY.
FT CONFLICT 1545 1545 T -> A (IN REF. 2).
SQ SEQUENCE 1742 AA; 193696 MW; 3F0699A2123EED07 CRC64;

Query Match 39.8%; Score 186.5; DB 1; Length 1742;
Best Local Similarity 42.5%; Pred. No. 1.4e-12;
Matches 37; Conservative 18; Mismatches 31; Indels 1; Gaps 1;

QY 2 SGGVKVQYKNDSPAGDNIKPGQLVNTGSSVDLSTVTYVFTDGGSTLVYNCNW 61
Db 704 SQGIKVLANKETNTTIRPWLKVNSSGSSIDLSTRTIRYWTVDGERAQSAIS-DW 762
QY 62 AAMGCGNIRASFGSVNPTPTADTYLQ 88
Db 763 AQIGASNVTFKVLSSSVSGADYILE 789

RESULT 6
GUNB_CALSA
ID GUNB_CALSA STANDARD; PRT; 1039 AA.
AC P10474;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Endoglucanase/exoglucanase B precursor [Includes: Endoglucanase
DE (EC 3.2.1.4) (Endo-1,4-beta-glucanase) (Cellulase)
DE (Cellobiohydrolase); Exoglucanase (EC 3.2.1.91) (Exocellobiohydrolase)
DE (1,4-beta-cellobiohydrolase)].
GN CELB.
OS Caldocellum saccharolyticum (Caldicellulosiruptor saccharolyticus).
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=44001;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89098398; PubMed=2789517;
RA Saul D.J., Williams L.C., Love D.R., Chamley I.W., Bergquist P.I.;
RT "Nucleotide sequence of a gene from Caldocellum saccharolyticum
RT encoding for exocellulase and endocellulase activity.";
RL Nucleic Acids Res. 17:439-439(1989).
CC -!- FUNCTION: THIS PROTEIN IS MADE UP OF TWO DOMAINS: THE N-TERMINAL
CC DOMAIN HAS EXOGLUCANASE ACTIVITY WHILE THE C-TERMINAL DOMAIN IS
CC AN ENDOGLUCANASE.
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC in cellulose and cellotetraose, releasing cellobiose from the non-
CC reducing ends of the chains.
CC -!- SIMILARITY: In the N-terminal section; belongs to cellulase family
CC F (family 10 of glycosyl hydrolases).
CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
CC A (FAMILY 5 OF GLYCOSYL HYDROLASES).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC modified and this statement is not removed. Usage by and for commercial

```

```
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X13602; CAA31936.1; -.
DR PIR; S02711; S02711.
DR HSSP; Q06851; INBC.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR001000; Glyco_hydro_10.
DR Pfam; PF00942; CEM_3; 1.
DR Pfam; PF00942; CEM_3; 1.
DR Pfam; PF00150; cellulase; 1.
DR Pfam; PF00331; Glyco_hydro_10; 1.
DR PRINTS; PR00134; GLHYDRLASE10.
DR ProDom; PD001947; CBD_3; 1.
DR SMART; SM00633; Glyco_10; 1.
DR PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR Cellulose degradation; Hydrolase; Glycosidase; Repeat;
KW Multifunctional enzyme; Signal.
FT SIGNAL 1 28
FT CHAIN 29 1039
FT DOMAIN 376 416
FT DOMAIN 417 570
FT DOMAIN 571 618
FT ACT_SITE 177 177
FT ACT_SITE 285 285
FT ACT_SITE 792 792
FT ACT_SITE 1039 1039
FT SEQUENCE 1039 AA; 117641 MW; 0E0378171594DDAE CRC64;
Query Match 38.9%; Score 182.5; DB 1; Length 1039;
Best Local Similarity 41.9%; Pred. No. 2.1e-12;
Matches 36; Conservative 18; Mismatches 31; Indels 1; Gaps 1;
QY 3 GGKVKVQKNDSPAGDNIKPKGLQVNTGSSVDLSTVTYVYFTRDGGSTLVNCDWA 62
DB 420 QIKVLVANKETSTTIRPWLKVNSSGSSIDLSRTVIRYVYTVDGERSAVS-DWA 478
QY 63 AMGCGNIRASFGSVNPATPTADTYLQ 88
DB 479 QIGASNTVTFKVLKSSVSGADYYLE 504
RESULT 7
YCEA_PAELA STANDARD; PRT; 145 AA.
AC P29718;
DT 01-APR-1993 (Rel. 25, Created)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein in cella 5' region (Fragment).
OS Paenibacillus lautus (Bacillus lautus).
OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.
OX NCBI_TaxID=1401;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PL236;
RX MEDLINE=92276330; PubMed=1592807;
RA Hansen C.K., Joergensen P.L., Diderichsen B.;
RT "cella from Bacillus lautus PL236 encodes a novel cellulose-binding
endo-beta-1,4-glucanase."
RL J. Bacteriol. 174:3522-3531(1992).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- PTM: A SHORT FORM (EGA-S) ARISES FROM POSTTRANSLATIONAL
PROTEOLYSIS OF APPROXIMATELY 150 AA AT THE C-TERMINUS OF EGA-L.
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY J (FAMILY 44 OF GLYCOSYL
HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M76588; AAA22303.1; -.
DR PIR; B41897; B41897.
DR HSSP; Q06851; INBC.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul_bind.
DR Pfam; PF00942; CEM_3; 1.
DR ProDom; PD001947; CBD_3; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 33
FT CHAIN 34 700
FT DOMAIN 548 700
FT ACT_SITE 213 213
FT SEQUENCE 700 AA; 76910 MW; 3D5C8CADA53EEB0F CRC64;
Query Match 36.7%; Score 172; DB 1; Length 700;
-----
```

```
DR HSSP; Q06851; INBC.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul_bind.
DR Pfam; PF00942; CEM_3; 1.
DR ProDom; PD001947; CBD_3; 1.
KW Hypothetical protein.
FT NON_TER 1
FT SEQUENCE 145 AA; 15782 MW; 9514E3A71B106AEB CRC64;
Query Match 38.5%; Score 180.5; DB 1; Length 145;
Best Local Similarity 43.9%; Pred. No. 3.4e-13;
Matches 36; Conservative 18; Mismatches 25; Indels 3; Gaps 2;
QY 7 VOYKNDSPAGDNIKPKGLQVNTGSSVDLSTVTYVYFTRDGGSTLVNCDWAAMGC 66
DB 1 LQYRAADTNAADNQIKPSFNKNKGTSVAVDLSTLKIRYFTKGSAAVNGW-IDWAQLGG 59
QY 67 GNIRASFGSVNPATPTADTYLQ 88
DB 60 SNIQISFG--NHTGNSDTYVE 79
RESULT 8
GUNA_PAELA STANDARD; PRT; 700 AA.
AC P29719;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase).
DE (Cellulase A) (EG-A).
GN CELA.
OS Paenibacillus lautus (Bacillus lautus).
OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.
OX NCBI_TaxID=1401;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PL236;
RX MEDLINE=92276330; PubMed=1592807;
RA Hansen C.K., Diderichsen B., Joergensen P.L.;
RT "cella from Bacillus lautus PL236 encodes a novel cellulose-binding
endo-beta-1,4-glucanase."
RL J. Bacteriol. 174:3522-3531(1992).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- PTM: A SHORT FORM (EGA-S) ARISES FROM POSTTRANSLATIONAL
PROTEOLYSIS OF APPROXIMATELY 150 AA AT THE C-TERMINUS OF EGA-L.
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY J (FAMILY 44 OF GLYCOSYL
HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M76588; AAA22303.1; -.
DR PIR; B41897; B41897.
DR HSSP; Q06851; INBC.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul_bind.
DR Pfam; PF00942; CEM_3; 1.
DR ProDom; PD001947; CBD_3; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 33
FT CHAIN 34 700
FT DOMAIN 548 700
FT ACT_SITE 213 213
FT SEQUENCE 700 AA; 76910 MW; 3D5C8CADA53EEB0F CRC64;
Query Match 36.7%; Score 172; DB 1; Length 700;
-----
```

**RESULT 9**

ID	GUNI_CLOTH	STANDARD;	PRT; 879 AA.
AC	Q02934;		
DT	01-FEB-1995 (Rel. 31, Created)		
DT	01-FEB-1995 (Rel. 31, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	Endoglucanase I precursor (EC 3.2.1.4) (EGI) (Endo-1,4-beta-glucanase) (Cellulase I).		
GN	Celi.		
OS	Clostridium thermocellum.		
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;		
OC	Clostridium.		
OX	NCBI_TaxID=1515;		
RN	[1]		
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 56-69.		
RC	STRAIN=NCIB 10682;		
RX	MEDLINE=93171873; PubMed=8436949;		
RA	Hazlewood G.P., Davidson K., Laurie J.I., Huskisson N.S., Gilbert H.J.;		
RT	"Gene sequence and properties of Celi, a family E endoglucanase from Clostridium thermocellum.";		
RL	J. Gen. Microbiol. 139:307-316(1993).		
CC	-!- FUNCTION: THIS ENZYME CATALYZES THE ENDOLYDROLYSIS OF 1,4-BETA-GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-GUANS. PRINCIPALLY ACTIVE AGAINST BARLEY BETA-GLUCAN.		
CC	-!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic linkages in cellulose, lichenin and cereal beta-D-glucans.		
CC	-!- PATHWAY: Cellulose degradation.		
CC	-!- SIMILARITY: Belongs to cellulase family E (family 9 of glycosyl hydrolases).		
-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).		
-----			
DR	EMBL; L04735; AAA20892.1; ..		
DR	FIR; A47704; A47704.		
DR	HSSP; P26221; LTFA.		
DR	InterPro; IPR001956; CBD 3.		
DR	InterPro; IPR008965; Cellul bind.		
DR	InterPro; IPR001701; Glyco_hydro_9.		
DR	InterPro; IPR008928; Glyco_trans_6hp.		
DR	Pfam; PF00942; CBM_3; 2.		
DR	Pfam; PF00759; Glyco_hydro_9; 1.		
DR	ProDom; PD001947; CBD 3; 1.		
DR	PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.		
DR	PROSITE; PS00592; GLYCOSYL_HYDROL_F9_2; 1.		
DR	PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.		
KW	Cellulose degradation; Hydrolase; Glycosidase; Signal.		
FT	SIGNAL 1 55		
FT	CHAIN 56 879		
FT	DOMAIN 56 518		
FT	DOMAIN 56 518		
FT	DOMAIN 729 879		
FT	ACT_SITE 448 448		
FT	BY SIMILARITY.		
FT	ACT_SITE 486 486		
FT	BY SIMILARITY.		
FT	ACT_SITE 495 495		
FT	BY SIMILARITY.		
SQ	SEQUENCE 879 AA; 97796 MW; 35A60069A514A927 CRC64;		

000 WORKING COPY FOR INFORMATION ONLY

```

FT DOMAIN 94 240 COHESIN 2.
FT DOMAIN 241 272 LINKER (PRO/THR-RICH).
FT DOMAIN 273 439 CELLULOSE-BINDING.
FT DOMAIN 440 461 LINKER (PRO/THR-RICH).
FT DOMAIN 462 607 COHESIN 3.
FT DOMAIN 710 733 DOCKERIN 1.
FT DOMAIN 743 766 DOCKERIN 2.
SQ SEQUENCE 772 AA; 82491 MW; BBF06DE5E094FE10 CRC64;

Query Match 35.2%; Score 165; DB 1; Length 772;
Best Local Similarity 36.1%; Pred. No. 1.4e-10;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKVOYKNDAPGNDQIKGLQLVNTGSSVDLSVTYVYWFTRDGGSSSTLVYVNC D 60
D 277 VSGNLKVEFYFNPSDDTNSINPQKVTNTGSSAIDLSKLTRYTYTVDGQKQDFW-CD 335
QY 61 WAAM-----CGGNIRASFGSVNPAATPTADTYLQ 88
D 336 HAAITGSGSYNGITSNVKGTGFKVMSSTNNADTYLE 372

RESULT 11
CIPA_CLOTM
ID CIPA_CLOTM STANDARD; PRT; 1853 AA.
AC Q06851;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cellulosomal scaffolding protein A precursor (Cellulosomal
DE glycoprotein SI/SL) (Cellulose integrating protein A) (Cohesin).
GN CIPA.
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1515;
RN [1]
RN SEQUENCE FROM N.A.A., AND SEQUENCE OF 29-40 AND 42-43.
RC STRAIN=ATCC 27405 / DSM 1237;
RX MEDLINE=93302508; PubMed=8316083;
RA Gerngross U.T., Romaniec M.P.M., Kobayashi T., Huskisson N.S.,
RA Demain A.L.;
RT "Sequencing of a Clostridium thermocellum gene (cipA) encoding the
RT cellulosomal SL-protein reveals an unusual degree of internal
RT homology.";
RL Mol. Microbiol. 8:325-334 (1993).
RN [2]
RN SEQUENCE OF 1820-1853 FROM N.A.
RX MEDLINE=93209931; PubMed=8458832;
RA Fujino T., Beguin P., Aubert J.-P.;
RT "Organization of a Clostridium thermocellum gene cluster encoding the
RT cellulosomal scaffolding protein Cipa and a protein possibly involved
RT in attachment of the cellulosome to the cell surface.";
RN J. Bacteriol. 175:1891-1899 (1993).
RN [3]
RX X-RAY CRYSTALLOGRAPHY (2.15 ANGSTROMS) OF 184-321.
RX MEDLINE=97238934; PubMed=9083107;
RA Shimon L.J.W., Bayer E.A., Morag E., Lamed R., Yaron S., Shoham Y.,
RA Frolow F.;
RT "A cohesin domain from Clostridium thermocellum: the crystal
RT structure provides new insights into cellulosome assembly.";
RL Structure 5:381-390 (1997).
RN [4]
RX X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 368-522.
RX MEDLINE=97076134; PubMed=8918451;
RA Tormo J., Lamed R., Chirino A.J., Morag E., Bayer E.A., Shoham Y.,
RA Steitz T.A.;
RT "Crystal structure of a bacterial family-III cellulose-binding
RT domain: a general mechanism for attachment to cellulose.";
RL EMBO J. 15:5739-5751 (1996).
RN [5]
RX X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 1213-1361.
RX MEDLINE=98022914; PubMed=9402065;

```

```

RA Tavares G.A., Beguin P., Alzari P.M.;
RT "The crystal structure of a type I cohesin domain at 1.7-A
RT resolution.";
RL J. Mol. Biol. 273:701-713 (1997).
CC 1- FUNCTION: ACTS AS A SCAFFOLDING PROTEIN IN THE CELLULOSE. IT
CC PROMOTES BINDING OF CELLULOSE TO THE CATALYTIC DOMAINS OF THE
CC CELLULOYTIC ENZYMES.
CC 1- SUBCELLULAR LOCATION: Cell surface.
CC 1- DOMAIN: THE COHESIN DOMAINS BIND TO THE DOCKERIN DOMAIN BORN BY
CC THE CATALYTIC COMPONENTS OF THE CELLULOSE.
CC 1- SIMILARITY: Contains 9 cohesin domains.
CC 1- SIMILARITY: Contains 2 dockerin domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC -----
CC EMBL; L08665; -; NOT ANNOTATED_CDS.
CC EMBL; X67506; CAA47840.1; -.
CC PIR; S36859; S36859.
CC PDB; 1ANU; 23-JUL-97.
CC PDB; 1AOC; 08-JUL-98.
CC PDB; 1NBC; 26-SEP-97.
CC InterPro; IPR001956; CBD 3.
CC InterPro; IPR008965; Cellul_bind.
CC InterPro; IPR002102; Cohesin.
CC InterPro; IPR002105; Dockerin_1.
CC InterPro; IPR002048; EF-hand.
CC Pfam; PF00942; CBM 3; 1.
CC Pfam; PF00963; Cohesin; 9.
CC Pfam; PF00404; Dockerin_1; 2.
CC ProDom; PD001947; CBD 3; 1.
CC PROSITE; PS00018; EF HAND; UNKNOWN 1.
CC PROSITE; PS00448; CLOS CELLULOSE_RPT; 2.
CC Cellulose degradation; Cell wall; Glycoprotein; Repeat; Signal;
CC 3D-structure.
FT CHAIN 1 28 CELLULOSONAL SCAFFOLDING PROTEIN A.
FT CHAIN 29 1853 COHESIN 1.
FT DOMAIN 29 182 COHESIN 2.
FT DOMAIN 183 322 COHESIN 3.
FT DOMAIN 323 353 LINKER (PRO/THR-RICH).
FT DOMAIN 364 522 CELLULOSE-BINDING (BY SIMILARITY).
FT DOMAIN 523 559 LINKER (PRO/THR-RICH).
FT DOMAIN 560 704 COHESIN 3.
FT DOMAIN 724 866 COHESIN 4.
FT DOMAIN 889 1031 COHESIN 5.
FT DOMAIN 1054 1196 COHESIN 6.
FT DOMAIN 1219 1361 COHESIN 7.
FT DOMAIN 1384 1526 COHESIN 8.
FT DOMAIN 1548 1690 COHESIN 9.
FT DOMAIN 1791 1814 DOCKERIN 1.
FT DOMAIN 1824 1847 DOCKERIN 2.
FT CONFLICT 1615 1615 A -> AA (IN REF. 1).
FT STRAND 185 188
FT STRAND 190 191
FT TURN 195 196
FT STRAND 198 206
FT TURN 210 211
FT STRAND 213 221
FT TURN 224 226
FT STRAND 227 234
FT TURN 236 237
FT TURN 243 246
FT TURN 247 252
FT TURN 253 256
FT STRAND 257 263
FT TURN 265 266
FT TURN 270 271
FT STRAND 273 273

```

```
FT STRAND 277 286
FT TURN 292 304
FT STRAND 305 306
FT STRAND 309 309
FT STRAND 313 315
FT STRAND 317 319
FT STRAND 369 375
FT STRAND 381 382
FT STRAND 385 385
FT STRAND 387 393
FT STRAND 399 400
FT HELIX 401 403
FT STRAND 404 410
FT STRAND 418 428
FT TURN 430 431
FT STRAND 434 436
FT HELIX 438 440
FT STRAND 441 452
FT TURN 453 454
FT STRAND 455 463
FT STRAND 467 468
FT TURN 470 471
FT STRAND 473 482
FT TURN 483 484
FT STRAND 488 489
FT TURN 491 492
FT TURN 494 495
FT STRAND 498 498
FT STRAND 503 504
FT STRAND 509 512
FT TURN 513 514
FT STRAND 515 518
FT STRAND 1220 1224
FT STRAND 1226 1229
FT TURN 1231 1232
FT STRAND 1234 1242
FT TURN 1246 1247
FT STRAND 1249 1249
FT STRAND 1251 1257
FT TURN 1260 1262
FT STRAND 1263 1270
FT TURN 1272 1273
FT HELIX 1279 1282
FT STRAND 1283 1288
FT TURN 1289 1292
FT STRAND 1293 1299
FT TURN 1301 1302
FT TURN 1306 1307
FT STRAND 1309 1309
FT STRAND 1313 1322
FT TURN 1324 1325
FT STRAND 1329 1342
FT TURN 1344 1345
FT STRAND 1348 1348
FT STRAND 1351 1354
FT STRAND 1356 1360
SQ SEQUENCE 1853 AA; 196831 MW; 3ABDDC03ABFC5372 CRC64;

Query Match 35.2%; Score 165; DB 1; Length 1853;
Best Local Similarity 36.1%; Pred. No. 3.9e-10;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKQVKNDSAPGDNQIKPGLQVNTGSSVDLSTVTYVYFTRDGGSTLYVNC 60
Db 365 VSGNLKVEFYNPSTTNSINPQFVYVNTGSAIDLSKILRYIYVYDQKDTFW-CD 423
QY 61 WAAM-----GCGNIRASFGSVNPTPTADTYLQ 88
Db 424 HAAIGSGSYNGITNSVKGTFVKMSSSTNNADTYLE 460

RESULT 12
GUNN_ERWCA
```

```
ID AC 059394;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Endoglucanase N precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase N)
DE (Cellulase N).
GN CELN.
OS Erwinia carotovora.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OX Enterobacteriaceae; Pectobacterium.
OX NCBI_TaxID=554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Atroseptica FCBR C18;
RX MEDLINE=9829944; PubMed=9636315;
RA Olsen O., Thomsen K.K., Weber J., Duus J.O., Svendsen I., Wegener C.,
RA von Wettstein D.;
RT "Transplanting two unique beta-glucanase catalytic activities into
RT one multienzyme, which forms glucose.";
RL Biotechnology 14:71-76(1996).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl
CC hydrolases).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L39788; AAC37033.1; --
DR HSSP; O85465; IA3H.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul bind.
DR InterPro; IPR001547; Glyco_Hydro_5.
DR Pfam; PF00942; CBM_3; 1.
DR Pfam; PF00150; cellulase; 1.
DR ProDom; PD001947; CBD_3; 1.
DR ProSITE; PS00659; GLYCOSYL HYDROL F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 31 POTENTIAL.
FT CHAIN 32 444 ENDOGLUCANASE N.
FT ACT_SITE 168 168 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 256 256 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 444 AA; 48300 MW; FA7E4179004CBB43 CRC64;

Query Match 31.9%; Score 149.5; DB 1; Length 444;
Best Local Similarity 38.6%; Pred. No. 3.8e-09;
Matches 34; Conservative 19; Mismatches 30; Indels 5; Gaps 2;

QY 2 SGGKVQYKNDSPGDNQIKPGLQVNTGSSVDLSTVTYVYFTRDGGSTLYVNC 59
Db 357 TGDVQLQYRVDNNPSDDAIRMAVNIKNTGSTPIKLSDLQVRYVYFHDGKGANLFV--- 413
QY 60 DWAMCGGNIRASFGSVNPTPTADTYL 87
Db 414 DWANVGPNNTVSTGTTPAASTDKANRYV 441

RESULT 13
GUNN_ERWCA
ID AC 047056;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Endoglucanase V precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase V)
DE (Cellulase V).
```



```

GN CELV.
OS Erwinia carotovora.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pectobacterium.
OX NCBI_TaxID=554;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=SCRI193;
RX MEDLINE=94067016; PubMed=8246888;
RA Cooper V.J.C., Salmond G.P.C.;
RT "Molecular analysis of the major cellulase (CelV) of Erwinia
RT carotovora: evidence for an evolutionary 'mix-and-match' of enzyme
RT domains.";
RL Mol. Gen. Genet. 241:341-350(1993).
CC -!- FUNCTION: Endoglucanase with some exoglucanase activity.
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MISCELLANEOUS: Has a pH optimum of about 7.0 and a temperature
CC optimum about 42 degrees Celsius.
CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl
CC hydrolases).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X76000; CAA53592.1; -.
DR PIR; S39962; S39962.
DR HSSP; O85465; 1A3H.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR008965; Cellul_bind.
DR Pfam; PF00942; CBM_3; 1.
DR ProDom; PD001947; Glyco_hydro_5.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR SIGNAL 1 31 POTENTIAL.
DR CHAIN 32 505 ENDOGLUCANASE V.
DR DOMAIN 32 334 CATALYTIC.
DR DOMAIN 335 352 LINKER.
DR DOMAIN 353 505 CELLULOSE-BINDING (BY SIMILARITY).
DR ACT_SITE 168 168 PROTON DONOR (BY SIMILARITY).
DR ACT_SITE 256 256 NUCLEOPHILE (BY SIMILARITY).
DR SEQUENCE 505 AA; 54900 MW; DBEA9337BB4D2623 CRC64;

Query Match 31.9%; Score 149.5; DB 1; Length 505;
Best Local Similarity 38.6%; Pred. No. 4.4e-09;
Matches 34; Conservative 19; Mismatches 30; Indels 5; Gaps 2;

QY 2 SGGVKQVKNDSAPGDNQIKPGLQVNTGSSVDLSTVTYRYWFTRDG--GSSTLVYNC 59
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
354 TGDVVLQYRNVNDNPSDDAIRMAVNIKNTGSTPIKLSDLQVRYFPHDDKPGANLFV--- 410

QY 60 DWAMGCGNIRASFGSVNPAATDTYL 87
Db ||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
411 DWANVGPNNTVSTGTTPAASTDKANRYV 438

RESULT 14
GUNW ERWCA
ID GUNW ERWCA STANDARD; PRT; 504 AA.
AC Q59395;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Endoglucanase V1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase V1)
(CELLULASE V1).

```

```

GN CELV1.
OS Erwinia carotovora.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pectobacterium.
OX NCBI_TaxID=554;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=SCC3193;
RX MEDLINE=95231512; PubMed=7715600;
RA Mae A., Heikinheimo R., Palva E.T.;
RT "Structure and regulation of the Erwinia carotovora subspecies
RT carotovora SCC3193 cellulase gene celV1 and the role of cellulase in
RT phytopathogenicity.";
RL Mol. Gen. Genet. 247:17-26(1995).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl
CC hydrolases).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X79241; CAA55823.1; -.
DR PIR; S54744; S54744.
DR HSSP; O85465; 1A3H.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR008965; Cellul_bind.
DR Pfam; PF00942; CBM_3; 1.
DR ProDom; PD001947; Glyco_hydro_5.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR SIGNAL 1 31 POTENTIAL.
DR CHAIN 32 504 ENDOGLUCANASE V1.
DR DOMAIN 32 334 CATALYTIC.
DR DOMAIN 335 352 LINKER.
DR DOMAIN 353 504 CELLULOSE-BINDING (BY SIMILARITY).
DR ACT_SITE 168 168 PROTON DONOR (BY SIMILARITY).
DR ACT_SITE 256 256 NUCLEOPHILE (BY SIMILARITY).
DR SEQUENCE 504 AA; 54963 MW; 0D7ECF74781565FA CRC64;

Query Match 31.4%; Score 147.5; DB 1; Length 504;
Best Local Similarity 37.5%; Pred. No. 7.4e-09;
Matches 33; Conservative 19; Mismatches 31; Indels 5; Gaps 2;

QY 2 SGGVKQVKNDSAPGDNQIKPGLQVNTGSSVDLSTVTYRYWFTRDG--GSSTLVYNC 59
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
354 TGDVVLQYRNVNDNPSDDAIRMAFNIKNTGSTPIKLSDLQVRYFPHDDKPGANLFV--- 410

QY 60 DWAMGCGNIRASFGSVNPAATDTYL 87
Db ||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
411 DWANVGPNNTVSTGTTPAASTDKANRYV 438

RESULT 15
GUN2 CLOSR
ID GUN2 CLOSR STANDARD; PRT; 914 AA.
AC P50900;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Exoglucanase II precursor (EC 3.2.1.91) (Exocellobiohydrolase II)
(1,4-beta-cellobiohydrolase II) (Avicelase II).
GN CELY.
OS Clostridium stercorarium.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;

```

```
OC Clostridium.
OX NCBI_TaxID=1510;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIB 11754;
RA Bronnenmeier K., Kundt K., Riedel K., Schwarz W.H.,
RA Staudenbauer W.L.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP CHARACTERIZATION.
RC STRAIN=NCIB 11754;
RC MEDLINE=91364586; PubMed=1909625;
RA Bronnenmeier K., Ruecknagel K.P., Staudenbauer W.L.;
RT "Purification and properties of a novel type of
RT exo-1,4-beta-glucanase (avicelase II) from the cellulolytic
RT thermophile Clostridium stercorarium.";
RL Eur. J. Biochem. 200:379-385(1991).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC in cellulose and celotetraose, releasing cellobiose from the non-
CC reducing ends of the chains.
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY L (FAMILY 48 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z69359; CAA93280.1; -.
DR HSSP; Q06851; INBC.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR005102; DUF291.
DR InterPro; IPR005556; Glyco_hydro_48.
DR InterPro; IPR008928; Glyco_trans_6hp.
DR InterPro; IPR007110; IG-like.
DR Pfam; PF00942; CBM 3; 1.
DR Pfam; PF03442; DUF291; 1.
DR Pfam; PF02011; Glyco_hydro_48; 1.
DR PRINTS; PR00844; GLHYDRLASE48.
DR ProDom; PD001947; CBD 3; 1.
DR ProDom; PD011903; Glyco_hydro_48; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 33 POTENTIAL.
FT CHAIN 34 914 EXOGLUCANASE II.
SQ SEQUENCE 914 AA; 103020 MW; D0DB6017D6DF82C CRC64;

Query Match 31.4%; Score 147.5; DB 1; Length 914;
Best Local Similarity 35.2%; Pred. No. 1.5e-08;
Matches 31; Conservative 21; Mismatches 35; Indels 1; Gaps 1;

QY 1 VSGGVKQVQKNDSAPGDNQIKPGQLVNTGSSVDLSTVTVYMFTRDGGSSTLVYNC 60
Db 763 VEGVLIQISFNANTQEISNIMPRFVYNSGNTSIPLSVKLYRYTVDDKPKQNF 821

QY 61 WAMCGNIRASFGSVNPATPTADTYLQ 88
Db 822 WASIGSSNVGTGTFVMDGATTGADYYLE 849
```

Search completed: May 11, 2004, 12:07:29  
Job time : 2.39913 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 11:45:55 ; Search time 10.2683 Seconds  
(without alignments)  
2704.020 Million cell updates/sec

Title: US-09-917-376-5

Perfect score: 469

Sequence: 1 VSGGVKQYKNDSAPGDNQ.....IRASFGSVNPTPTADTYLVQ 88

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_25:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phase:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_virus:\*

16: sp\_bacteriap:\*

17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query %	Match	Length	DB	ID	Description
1	257	54.8	741	16	Q82QF2		Q82Qf2 streptomyc
2	228.5	48.7	616	2	Q7X2N2		Q7x2n2 thermomonos
3	225.5	48.1	170	2	Q9RFK6		Q9rfx6 caldibacill
4	216.5	46.2	930	2	Q9RFK5		Q9rfx5 caldibacill
5	214.5	45.7	921	2	Q9L8L8		Q9l8l8 caldibacill
6	209.5	44.7	1091	2	Q8KKF7		Q8kkf7 paenibacill
7	204.5	43.6	997	2	Q9Z4I1		Q9z4i1 bacillus sp
8	194.5	41.5	1751	2	Q9AQG4		Q9agg4 caldicellul
9	192.5	41.0	1000	2	Q24820		Q24820 thermophili
10	192.5	41.0	1770	2	Q9X3P5		Q9x3p5 caldicellul
11	191.5	40.8	261	2	Q9AQG7		Q9agg7 caldicellul
12	191.5	40.8	1426	2	Q9X3P6		Q9x3p6 caldicellul
13	191	40.7	1711	2	P96311		P96311 anaerocellul
14	190.5	40.6	996	2	Q9AQH0		Q9aqh0 caldicellul
15	190.5	40.6	1779	2	O52374		O52374 caldicellul
16	190	40.5	499	2	Q93TJ6		Q93tj6 bacillus su

17	190	40.5	508	2	O93LD0		O93ld0 bacillus su
18	187	39.9	499	2	O52731		O52731 bacillus sp
19	186	39.7	499	2	O45532		O45532 bacillus su
20	186	39.7	501	2	O83012		O83012 bacillus sp
21	180	38.4	486	2	O45430		O45430 bacillus sp
22	177	37.7	499	2	O8RPO6		O8rpo6 bacillus am
23	172.5	36.8	1915	2	O9RPL0		O9rpl0 acetivibrio
24	168	35.8	473	16	Q9RK75		Q9rk75 streptomyc
25	165.5	35.3	887	2	Q9L3J8		Q9l3j8 clostridium
26	157	33.5	542	2	Q7X3S6		Q7x3s6 bacillus li
27	150.5	32.1	2316	2	Q9FDJ9		Q9fdj9 bacteroides
28	130.5	27.8	1483	16	Q977Y4		Q977y4 clostridium
29	130.5	27.8	1546	2	Q45996		Q45996 clostridium
30	126.5	27.0	1162	2	O82830		O82830 clostridium
31	119.5	25.5	307	2	Q46392		Q46392 clostridium
32	117	24.9	1230	2	O59325		O59325 clostridium
33	107	22.8	221	2	O8VVI7		O8vvi7 clostridium
34	87	18.6	440	2	Q93LI9		Q93li9 bacillus sp
35	78	16.6	135	13	Q90542		Q90542 ginglymosto
36	76	16.2	121	13	Q8JGA5		O8jga5 ginglymosto
37	76	16.2	121	13	Q8JGA9		O8jga9 ginglymosto
38	75.5	16.1	618	4	Q9Y3Z2		Q9y3z2 homo sapien
39	75.5	16.1	660	2	O9L3J2		O9l3j2 clostridium
40	75.5	16.1	986	4	O94858		O94858 homo sapien
41	75.5	16.1	1300	4	O8IZF2		O8izf2 homo sapien
42	73.5	15.7	547	5	P91006		P91006 caenorhabdi
43	73.5	15.7	1428	16	Q8YRU7		Q8yru7 anabaena sp
44	73	15.6	511	5	Q9GVG5		Q9gvg5 caenorhabdi
45	72	15.4	499	12	Q993M3		Q993m3 autonomic

#### ALIGNMENTS

#### RESULT 1

Q82QF2	Q82QF2	PRELIMINARY;	PRT;	741 AA.
AC	Q82QF2;			
DT	01-JUN-2003 (TREMELrel. 24, Created)			
DT	01-JUN-2003 (TREMELrel. 24, Last sequence update)			
DT	01-OCT-2003 (TREMELrel. 25, Last annotation update)			
DE	Putative cellulose 1,4-beta-cellobiosidase.			
GS	GUXAI OR SAV557.			
OS	Streptomyces avermitilis.			
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OX	Streptomycineae; Streptomycetaceae; Streptomyces.			
OX	NCBI_TaxID=33903;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=WA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;			
RX	MEDLINE=21477403; PubMed=11572948;			
RA	Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,			
RA	Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,			
RA	Kikuchi H., Shiba T., Sakaki Y., Hattori M.;			
RT	"Genome sequence of an industrial microorganism Streptomyces			
RT	avermitilis: deducing the ability of producing secondary			
RT	metabolites.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=WA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;			
RX	MEDLINE=22608306; PubMed=12692562;			
RA	Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,			
RA	Sakaki Y., Hattori M., Omura S.;			
RT	"Complete genome sequence and comparative analysis of the industrial			
RT	microorganism Streptomyces avermitilis.";			
RL	Nat. Biotechnol. 21:526-531(2003).			
DR	EMBL; AP005023; BAC68267.1; -			
DR	GO; GO:0004553; Fhydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.			
DR	GO; GO:0005975; P:carbohydrate metabolism; IEA.			
DR	InterPro; IPR001956; CBD 3.			
DR	InterPro; IPR008965; Cellul. bind.			
DR	InterPro; IPR003961; FN_III.			

```
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR001524; Glyco_hydro_6.
DR Pfam; PF00942; CEM_3; 1.
DR Pfam; PF00041; fn3; 1.
DR Pfam; PF01341; Glyco_hydro_6; 1.
DR PRINTS; PR00733; GLYDRLAS6.
DR ProDom; PD001947; CBD_3; 1.
DR ProDom; PD003733; Glyco_hydro_6; 1.
DR SMART; SM00060; FN3; 1.
DR PROSITE; PS00655; GLYCOSYL_HYDROL_F6_1; 1.
KW Complete proteome.
SQ SEQUENCE 741 AA; 77396 MW; 79404B40B2B4A7AF CRC64;

Query Match 54.8%; Score 257; DB 16; Length 741;
Best Local Similarity 56.3%; Pred. No. 1.7e-19;
Matches 49; Conservative 13; Mismatches 25; Indels 0; Gaps 0;

QY 2 SGVKVQYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTYVWFTRDGGSTLVNCDW 61
Db 591 SGGLKVLKNDSSATDNRAIRPLRVNTGSGSLDSKVTRYFTSDSGSPTVNAWCDY 650
QY 62 AAMGCNIRASFGSVNPTPTADTYLQ 88
Db 651 AAVGCSVSLKVPLTTPVPGADAYLE 677

RESULT 2
QYX2N2 PRELIMINARY; PRT; 616 AA.
AC Q7X2N2;
DT 01-OCT-2003 (TREMBlrel. 25, Created)
DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Endoglucanase.
GN CEL5B.
OS Thermomonospora fusca.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptosporangineae; Nocardiopsaceae; Thermobifida.
OX NCBI_TaxID=2021;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TM51;
RA Posta K., Beki E., Kukolya J., Hornok L.;
RT "Phylogenetic relationships of Tt cel5B, a new endoglucanase encoding
  gene from Thermobifida fusca.";
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY298814; AAP56348.1; -.
SQ SEQUENCE 616 AA; 67701 MW; 24FFC1EAI3F5639 CRC64;

Query Match 48.7%; Score 228.5; DB 2; Length 616;
Best Local Similarity 52.3%; Pred. No. 1.8e-16;
Matches 45; Conservative 16; Mismatches 22; Indels 3; Gaps 2;

QY 2 SGVKVQYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTYVWFTRDGGSTLVNCDW 61
Db 471 TGALEVVYRNNSLAADDSQIAPGLRVNTGSSVDLADVEIHYYFTNEPG-GTLQFTCDW 529
QY 62 AAMGCNIRASFGSVNPTPTADTYL 87
Db 530 AQVGCANVNASFTSL-SAPGADTSL 553

RESULT 3
QYRFK6 PRELIMINARY; PRT; 170 AA.
AC QYRFK6;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein (fragment).
OS Caldicellus cellulovorans.
OC Bacteria; Firmicutes; Bacillales; Alicyclobacillaceae; Caldicellus.
OX NCBI_TaxID=74586;
```

```

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20120520; PubMed=10653733;
RA Summa A., Gibbs M.D., Chin C.W.J., Nelson P.J., Bergquist P.L.;
RT "A gene encoding a novel multidomain beta-1,4-mannanase from
  Caldicellus cellulovorans and action of the recombinant enzyme on
  kraft pulp.";
RL Appl. Environ. Microbiol. 66:664-670(2000).
DR HMBL; AF163837; AAF22273.1; -.
DR HSSP; Q06851; INBC.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul_bind.
DR Pfam; PF00942; CEM_3; 1.
DR ProDom; PD001947; CBD_3; 1.
KW Hypothetical protein_1
FT NON_TER 1
SQ SEQUENCE 170 AA; 18493 MW; 7AC9D33F44E3A0B4 CRC64;

Query Match 48.1%; Score 225.5; DB 2; Length 170;
Best Local Similarity 47.7%; Pred. No. 8.2e-17;
Matches 41; Conservative 17; Mismatches 27; Indels 1; Gaps 1;

QY 3 GGVKQYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTYVWFTRDGGSTLVNCDW 62
Db 20 GSLVVQYRAADTNAGDNQIKPFRIVNRGTSSVPLSELITRYWYTVD-GDKPQVFNCDWA 78
QY 63 AMGCNIRASFGSVNPTPTADTYLQ 88
Db 79 QVGCNVRGSEFVKLSTGRTGADYYIE 104

RESULT 4
QYRFK5 PRELIMINARY; PRT; 930 AA.
AC QYRFK5;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Multidomain beta-1,4-mannanase precursor.
GN MANA.
OS Caldicellus cellulovorans.
OC Bacteria; Firmicutes; Bacillales; Alicyclobacillaceae; Caldicellus.
OX NCBI_TaxID=74586;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20120520; PubMed=10653733;
RA Summa A., Gibbs M.D., Chin C.W.J., Nelson P.J., Bergquist P.L.;
RT "A gene encoding a novel multidomain beta-1,4-mannanase from
  Caldicellus cellulovorans and action of the recombinant enzyme on
  kraft pulp.";
RL Appl. Environ. Microbiol. 66:664-670(2000).
DR HMBL; AF163837; AAF22274.1; -.
DR HSSP; Q06851; INBC.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR004302; Chitin_binding_3.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00942; CEM_3; 2.
DR Pfam; PF00150; cellulase; 1.
DR Pfam; PF03067; Chitin_bind_3; 1.
DR ProDom; PD001947; CBD_3; 2.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW SIGNAL.
FT CHAIN 1 33 POTENTIAL
FT CHAIN 34 930 MULTIDOMAIN BETA-1,4-MANNANASE.
SQ SEQUENCE 930 AA; 101576 MW; 0086638D54D1A2CC CRC64;

Query Match 46.2%; Score 216.5; DB 2; Length 930;
Best Local Similarity 46.5%; Pred. No. 6.2e-15;
```

```
Matches 40; Conservative 17; Mismatches 28; Indels 1; Gaps 1;
QY 3 GGVKQVKNDSAPGDNIQKGLQLVNTGSSVDLSTVTYRVYFTRDGGSSFLVYNCWA 62
   :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
Db 780 GNLVVQVRAADTNATNQLKPHFRIVNRGTSSVPLSELTIRYWTVD-GDKPQVNCWA 838
QY 63 AMGCCNIRASFGSNVPATPTADTYLQ 88
   :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
Db 839 QVGCNLRGSLFKLSTGRGTADYVIE 864

RESULT 5
Q9L8L8
ID Q9L8L8 PRELIMINARY; PRT; 921 AA.
AC Q9L8L8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Beta-1,4-xylanase Xyna precursor.
GN Xyna.
OS Caldicabacillus cellulosovorans.
OC Bacteria; Firmicutes; Bacillales; Alicyclobacillaceae; Caldicabacillus.
OX NCBI_TaxID=74586;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20519260; PubMed=11065373;
RA Sunna A., Gibbs M.D., Bergquist P.L.;
RT "A novel thermostable multidomain 1,4-beta-xylanase from
RT 'Caldicabacillus cellulosovorans' and effect of its xylan-binding domain
RT on enzyme activity.";
RL Microbiology 146:2947-2955 (2000).
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY F (FAMILY 10 OF GLYCOSYL
CC HYDROLASES).
DR EMBL; AF200304; AAF61649.1; -.
DR HSSP; Q06851; INCB.
DR GO; GO:0004553; P:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR003305; CBM_CenC.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR008979; Gal_bind_like.
DR InterPro; IPR001000; Glyco_hydro_10.
DR Pfam; PF00942; CBM 3; 2.
DR Pfam; PF02018; CBM 4; 9; 1.
DR Pfam; PF00331; Glyco_hydro_10; 1.
DR PRINTS; PR00134; GLYDRLASE10.
DR ProDom; PD001947; CBD_3; 2.
DR SMART; SM00633; Glyco_10; 1.
DR Glycosidase; Hydrolase; Signal; Xylan degradation.
KW SIGNAL
FT CHAIN 1 33 POTENTIAL.
SQ SEQUENCE 921 AA; 102380 MW; C5DDDLIAF7567413 CRC64;

Query Match 45.7%; Score 214.5; DB 2; Length 921;
Best Local Similarity 46.5%; Pred. No. 1e-14;
Matches 40; Conservative 16; Mismatches 29; Indels 1; Gaps 1;

QY 3 GGVKQVKNDSAPGDNIQKGLQLVNTGSSVDLSTVTYRVYFTRDGGSSFLVYNCWA 62
   :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
Db 771 GNLVVQVRAADTNAGDNQLKPHFRIVNRGTSSVPLSELTIRYWTVD-GDKPQVNCWA 829
QY 63 AMGCCNIRASFGSNVPATPTADTYLQ 88
   :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
Db 830 WGCNLRGSLVKLTTGRGTADYVIE 855

RESULT 6
Q8KKF7
ID Q8KKF7 PRELIMINARY; PRT; 1091 AA.
AC Q8KKF7;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
```

```
DE Cellulose 1,4-beta-cellobiosidase precursor (EC 3.2.1.1.91).
GN CEL48C.
OS Paenibacillus sp. BP-23.
OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.
OX NCBI_TaxID=198119;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP-23;
RA Sanchez M.M., Pastor F.I.J., Diaz P.;
RT "Paenibacillus sp. BP-23 family 48 cellulase. Cloning and performance
RT on cellulosic substrates.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
DR EMBL; AJ488933; CAD32945.1; -.
DR GO; GO:0016162; F:cellulose 1,4-beta-cellobiosidase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR000556; Glyco_hydro_48.
DR InterPro; IPR008928; Glyco_trans_6hp.
DR Pfam; PF00942; CBM 3; 1.
DR Pfam; PF00641; fn3; 1.
DR Pfam; PF02011; Glyco_hydro_48; 1.
DR PRINTS; PR00844; GLHYDRLASE48.
DR ProDom; PD001947; CBD 3; 1.
DR ProDom; PD011903; Glyco_hydro_48; 1.
DR SMART; SM00060; FN3; 2.
KW Signal; Hydrolase; Glycosidase.
FT SIGNAL 1 35 POTENTIAL.
FT CHAIN 36 1091 CELLULOSE 1,4-BETA-CELLOBIOSIDASE.
SQ SEQUENCE 1091 AA; 118001 MW; 21EACCEB2E704478 CRC64;

Query Match 44.7%; Score 209.5; DB 2; Length 1091;
Best Local Similarity 46.0%; Pred. No. 4.4e-14;
Matches 40; Conservative 21; Mismatches 25; Indels 1; Gaps 1;

QY 2 SGGVKQVKNDSAPGDNIQKGLQLVNTGSSVDLSTVTYRVYFTRDGGSSFLVYNCW 61
   :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
Db 940 TGLEVOYRNGSSASGNAITPQNLKNTGTTAIDLSKVKRYFTKD-SAADMSFWCDY 998
QY 62 AMGCCNIRASFGSNVPATPTADTYLQ 88
   :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
Db 999 AQLGSANVQGSFVAVNPAKGTADTYLE 1025

RESULT 7
Q9Z4I1
ID Q9Z4I1 PRELIMINARY; PRT; 997 AA.
AC Q9Z4I1;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Cellulase precursor (EC 3.2.1.4).
GN CELB.
OS Bacillus sp. BP-23.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=89769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP-23;
RX MEDLINE=21129642; PubMed=11234960;
RA Pastor F.I.J., Pujol X., Blanco A., Vidal T., Torres A.L., Diaz P.;
RT "Molecular cloning and characterization of a multidomain endoglucanase
RT from Paenibacillus sp BP-23: evaluation of its performance in pulp
RT refining.";
RL Appl. Microbiol. Biotechnol. 55:61-68 (2001).
CC -!- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
DR EMBL; AJ133614; CAB38941.1; -.
DR HSSP; P26221; 1TF4.
DR GO; GO:0008810; F:cellulase activity; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
```

```
DR GO: GO:0005975; P:carbohydrate metabolism; IEA.
DR GO: GO:0006810; P:transport; IEA.
DR InterPro: IPR001956; CBD 3.
DR InterPro: IPR008965; Cellul bind.
DR InterPro: IPR003961; FN III-like.
DR InterPro: IPR008957; FN III-like.
DR InterPro: IPR001701; Glyco_hydro_9.
DR InterPro: IPR008928; Glyco_trans_6hp.
DR InterPro: IPR000566; Lipocin_cytFABP.
DR Pfam: PF00942; CBM 3; 2.
DR Pfam: PF00041; fn3_2.
DR Pfam: PF00759; Glyco_hydro_9; 1.
DR ProDom: PD001947; CBD 3; 1.
DR SMART: SM00060; FN3; 2.
DR PROSITE: PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE: PS00698; GLYCOSYL_HYDROL_F9_2; 1.
DR PROSITE: PS00213; LIPOCALIN; 1.
KW Glycosidase; Hydrolase; Signal.
FT SIGNAL 1 34 POTENTIAL.
SQ SEQUENCE 997 AA; 106927 MW; F20CB119D6410629 CRC64;

Query Match 43.6%; Score 204.5; DB 2; Length 997;
Best Local Similarity 43.7%; Pred. No. 1.4e-13;
Matches 38; Conservative 21; Mismatches 27; Indels 1; Gaps 1;

QY 2 SGGVKVQYKNDSDAPGDNQIKFGLQVNTGSSVDLSTVTYVWFTRDGSSSTLVNCDW 61
Db 846 TGTLEVQYRSGSGNSNAVTQFNLKNTGTQADLSTVKIRYFTKO-GTEELSFWDY 904

QY 62 AAMGCGNIRASFGSVNPATPTADTYLQ 88
Db 905 AQVGSANVQGMFVAVNPAKGTADTYVE 931

RESULT 8
Q9AQA4 PRELIMINARY; PRT; 1751 AA.
AC Q9AQA4;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE CelE.
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tok7B.1;
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
RT thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL Curr. Microbiol. 40:333-340 (2000).
DR EMBL; AF078042; AAK06394.1; -.
DR HSSP; Q06851; INBC.
DR GO: GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO: GO:0005215; P:transporter activity; IEA.
DR GO: GO:0005975; P:carbohydrate metabolism; IEA.
DR GO: GO:0006810; P:transport; IEA.
DR InterPro: IPR001956; CBD 3.
DR InterPro: IPR008965; Cellul bind.
DR InterPro: IPR001701; Glyco_hydro_9.
DR InterPro: IPR008928; Glyco_trans_6hp.
DR InterPro: IPR000566; Lipocin_cytFABP.
DR Pfam: PF00942; CBM 3; 4.
DR Pfam: PF00759; Glyco_hydro_9; 1.
DR ProDom: PD001947; CBD 3; 3.
DR PROSITE: PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE: PS00698; GLYCOSYL_HYDROL_F9_2; 1.
DR PROSITE: PS00213; LIPOCALIN; 1.
SQ SEQUENCE 1751 AA; 192176 MW; 60178CBF3C00BE95 CRC64;

Query Match 41.0%; Score 192.5; DB 2; Length 1000;
Best Local Similarity 43.7%; Pred. No. 2.9e-12;
Matches 38; Conservative 17; Mismatches 31; Indels 1; Gaps 1;

QY 2 SGGVKVQYKNDSDAPGDNQIKFGLQVNTGSSVDLSTVTYVWFTRDGSSSTLVNCDW 61
Db 372 SGQIKVLYANKETNTSTIRPWLKVNTGSSSIDLSRVIRYVWTVVDGDKAQSAVS-DW 430

QY 62 AAMGCGNIRASFGSVNPATPTADTYLQ 88
Db 431 AQIGASNVTFKFKVLSSSVSGADYILE 457

RESULT 10
Q9X3P5 PRELIMINARY; PRT; 1770 AA.
AC Q9X3P5;
ID Q9X3P5;
SQ SEQUENCE 1770 AA; 192176 MW; 60178CBF3C00BE95 CRC64;
```

```

DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE KYNA.
GN XNA.
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tok7B.1.
RX MEDLINE=20171169; PubMed=10706665;
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
RL thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL Curr. Microbiol. 40:333-340(2000).
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY F (FAMILY 10 OF GLYCOSYL
CC HYDROLASES).
DR EMBL; AF078737; AAD30363.1; -.
DR HSSP; Q06851; INBC.
DR GO; GO:0030246; F:carbohydrate binding; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR006584; CBD IV.
DR InterPro; IPR005084; CBM_6.
DR InterPro; IPR003305; CBM_CenC.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR008979; Gal_bind_like.
DR InterPro; IPR001000; Glyco_hydro_10.
DR Pfam; PF00942; CBM 3; 3.
DR Pfam; PF02018; CBM 4; 9; 2.
DR Pfam; PF03422; CBM_6; 1.
DR Pfam; PF00331; Glyco_hydro_10; 1.
DR PRINTS; PRO0134; GLHYDRLASE10.
DR ProDom; PD001947; CBD 3; 3.
DR SMART; SM00606; CBD IV; 1.
DR SMART; SM00633; Glyco_10; 1.
DR PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 1770 AA; 193641 MW; 8BAF1937D4926C92 CRC64;

Query Match 41.0%; Score 192.5; DB 2; Length 1770;
Best Local Similarity 43.5%; Pred. No. 5.7e-12;
Matches 37; Conservative 16; Mismatches 31; Indels 1; Gaps 1;

QY 4 GVKVQYKNDSPAGDNQIKPGLQVNTGSSVDLSVTYVRYWFTRDGSSSTLVNCDWAA 63
DB 1104 GLKVLKNNETSASTSSIRPFWFKIVNGSSVDLSRVKIRYWTVDGDKPOS AV-CDWAAQ 1162
QY 64 MCGGNIRASFGSVNTPATPTADTYLQ 88
DB 1163 IGASNVTFNFVKLSGSGVADYYLE 1187

RESULT 11
Q9A0G7 PRELIMINARY; PRT; 261 AA.
AC Q9A0G7;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Glycosyl hydrolase 6 (Fragment)
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tok7B.1.
RX MEDLINE=20171169; PubMed=10706665;

```

```

RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
RL thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL Curr. Microbiol. 40:333-340(2000).
DR EMBL; AF078040; AAK06391.1; -.
DR HSSP; Q06851; INBC.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR008965; Cellul_bind.
DR Pfam; PF00942; CBM 3; 1.
DR ProDom; PD001947; CBD_3; 1.
KW Hydrolase.
FT NON_TER 261
SQ SEQUENCE 261 AA; 28759 MW; 4771744A26A6AE04 CRC64;

Query Match 40.8%; Score 191.5; DB 2; Length 261;
Best Local Similarity 43.5%; Pred. No. 7.3e-13;
Matches 37; Conservative 15; Mismatches 32; Indels 1; Gaps 1;

QY 4 GVKVQYKNDSPAGDNQIKPGLQVNTGSSVDLSVTYVRYWFTRDGSSSTLVNCDWAA 63
DB 118 GLKVLKNNETSASTSSIRPFWFKIVNGSSVDLSRVKIRYWTVDGDKPOS AV-CDWAAQ 176
QY 64 MCGGNIRASFGSVNTPATPTADTYLQ 88
DB 177 IGASNVTFNFVKLSGSGVADYYLE 201

RESULT 12
Q9X3P6 PRELIMINARY; PRT; 1426 AA.
ID Q9X3P6;
AC Q9X3P6;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CELB.
GN CELB.
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tok7B.1.
RX MEDLINE=20171169; PubMed=10706665;
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
RL thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL Curr. Microbiol. 40:333-340(2000).
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY F (FAMILY 10 OF GLYCOSYL
CC HYDROLASES).
DR EMBL; AF078737; AAD30364.1; -.
DR HSSP; Q06851; INBC.
DR GO; GO:0003779; F:actin binding; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001589; Actbind_actnin.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR001000; Glyco_hydro_10.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00942; CBM 3; 3.
DR Pfam; PF00150; cellulase; 1.
DR Pfam; PF00331; Glyco_hydro_10; 1.
DR PRINTS; PRO0134; GLHYDRLASE10.
DR ProDom; PD001947; CBD_3; 3.
DR SMART; SM00633; Glyco_10; 1.
DR PROSITE; PS00019; ACTININ 1; 1.
DR PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.

```

```
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 1426 AA; 157544 MW; 29B3FDB85D09A863 CRC64;

Query Match 40.8%; Score 191.5; DB 2; Length 1426;
Best Local Similarity 43.5%; Pred. No. 5.7e-12;
Matches 37; Conservative 15; Mismatches 32; Indels 1; Gaps 1;

QY 4 GVKVOYKNDSPAGDNQIKPGLQLVNTGSSVDLSVTYRVYFTRDGGSSSTLVNCDWAA 63
Db 413 GLKVLVKNNETSASTGSIREFWFKVNGSSVDLSRVKIRYWTVDGDKPOS AV-CDWAQ 471

QY 64 MCGGNIRASFGSVNPATPTADTYLQ 88
Db 472 IGASNVTFNFVKLSGSGVADYILE 496

RESULT 13
ID P96311 PRELIMINARY; PRT; 1711 AA.
AC P96311;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE 1,4-beta-glucanase (fragment).
GN CELA.
OS Anaerocellum thermophilum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Anaerocellum group;
OC Anaerocellum.
OX NCBI_TaxID=31899;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z-1320;
RX MEDLINE=98154434; PubMed=9493383;
RA Zverlov V., Mahr S., Riedel K., Bronnenmeier K.;
RT "Properties and gene structure of a bifunctional cellulolytic enzyme
RT (Cela) from the extreme thermophile Anaerocellum thermophilum with
RT separate glycosyl hydrolase family 9 and 48 catalytic domains.";
RL Microbiology 144:457-465(1998).
DR EMBL; Z86105; CAB06786.1; -.
DR PIR; T31337; T31337.
DR HSSP; P26221; ITF4.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR008965; Cellul bind.
DR InterPro; IPR000556; Glyco_hydro_48.
DR InterPro; IPR001701; Glyco_hydro_9.
DR InterPro; IPR008928; Glyco_trans_6hp.
DR Pfam; PF00942; CBM_3; 3.
DR Pfam; PF02011; Glyco_hydro_48; 1.
DR Pfam; PF00759; Glyco_hydro_9; 1.
DR PRINTS; PR00844; GLHYDRLASE48.
DR ProDom; PD001947; CBD_3; 2.
DR ProDom; PD011903; Glyco_hydro_48; 1.
DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.
FT NON TER 1
FT SEQUENCE 1711 AA; 189979 MW; E3E987CEB9CD0C21 CRC64;

Query Match 40.7%; Score 191; DB 2; Length 1711;
Best Local Similarity 43.8%; Pred. No. 8e-12;
Matches 39; Conservative 18; Mismatches 30; Indels 2; Gaps 2;

QY 1 VSGG-KVVOYKNDSPAGDNQIKPGLQLVNTGSSVDLSVTYRVYFTRDGGSSSTLVNCDWAA 59
Db 683 VAGGQIKVLVANKNETSTTIRPWLKVNTGSSVDLSRVKIRYWTVDGDKPOS AV-741

QY 60 DWAMCGGNIRASFGSVNPATPTADTYLQ 88
Db 742 DWAGIGASNVTFKFKVLSGSGVADYILE 770

RESULT 14
ID Q9AQHO PRELIMINARY; PRT; 996 AA.
AC Q9AQHO;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Glycosyl hydrolase 5 (fragment).
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tok7B.1;
RX MEDLINE=20171169; PubMed=10706665;
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
RT thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL Curr. Microbiol. 40:333-340(2000).
DR EMBL; AF078038; AAK06388.1; -.
DR HSSP; Q06851; INBC.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR008965; Cellul bind.
DR InterPro; IPR002860; GH_BNR.
DR Pfam; PF02012; BNR_9.
DR Pfam; PF00942; CBM_3; 1.
DR ProDom; PD001947; CBD_3; 1.
DR KW Hydrolase.
DR NON TER 996
FT SEQUENCE 996 AA; 108275 MW; 3C72B6ED22E3C614 CRC64;

Query Match 40.6%; Score 190.5; DB 2; Length 996;
Best Local Similarity 43.5%; Pred. No. 4.8e-12;
Matches 37; Conservative 15; Mismatches 32; Indels 1; Gaps 1;

QY 4 GVKVOYKNDSPAGDNQIKPGLQLVNTGSSVDLSVTYRVYFTRDGGSSSTLVNCDWAA 63
Db 844 GLKVLVKNNETSASTGSIREFWFKVNGSSVDLSRVKIRYWTVDGDKPOS AV-CDWAQ 902

QY 64 MCGGNIRASFGSVNPATPTADTYLQ 88
Db 903 IGASNVTFNFVKLSGSGVADYILE 927

RESULT 15
ID O52374 PRELIMINARY; PRT; 1779 AA.
AC O52374;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Family 10 xylanase (EC 3.2.1.8).
GN XNC.
OS Caldicellulosiruptor sp. Rt69B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=70295;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rt69B.1;
RA Morris D.D., Gibbs M.D., Ford M., Thomas J., Bergquist P.L.;
RA "Family 10 and 11 xylanase genes from Caldicellulosiruptor sp.
RT Rt69B.1.";
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY F (FAMILY 10 OF GLYCOSYL
CC HYDROLASES).
CC EMBL; AF036924; AAB95326.1; -.
DR PIR; T31085; T31085.
DR HSSP; Q06851; INBC.
```



Search completed: May 11, 2004, 12:09:29  
Job time : 11.2683 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 12:04:16 ; Search time 4.31843 Seconds  
(without alignments)  
1052.023 Million cell updates/sec

Title: US-09-917-376-5

Perfect score: 469

Sequence: 1 VSGGVKQYKXNDSPAGDNQ.....IRASFGSVNPTPTADTYLQ 88

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.\*

- 1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/2/iaa/PTCUS\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	195.5	41.7	616	3	US-09-136-574A-47
2	194.5	41.5	1751	3	US-09-136-574A-44
3	191.5	40.8	1426	3	US-09-136-574A-43
4	172	36.7	700	2	US-07-862-588B-2
5	165.5	35.3	551	2	US-09-033-537A-1
6	165	35.2	167	5	PCT-US95-13813-9
7	165	35.2	476	4	US-09-339-159B-4
8	165	35.2	493	3	US-09-198-956-10
9	165	35.2	493	3	US-09-198-955A-12
10	165	35.2	493	4	US-09-694-531-12
11	165	35.2	493	4	US-09-670-141-10
12	165	35.2	493	4	US-10-072-152-12
13	120.5	25.7	531	2	US-07-862-588B-7
14	114	24.3	162	1	US-08-048-164A-2
15	114	24.3	162	1	US-08-460-462-2
16	114	24.3	162	1	US-08-460-457-2
17	114	24.3	162	1	US-08-460-458-2
18	114	24.3	162	2	US-08-460-455-2
19	114	24.3	162	2	US-08-330-394A-2
20	114	24.3	163	3	US-09-006-636-7
21	114	24.3	163	3	US-09-006-632-7
22	114	24.3	163	4	US-09-325-274-7
23	113	24.1	382	3	US-09-277-716-22
24	113	24.1	382	4	US-09-609-161B-22
25	112	23.9	154	2	US-08-330-394A-29
26	112	23.9	156	2	US-08-330-394A-22
27	64	13.6	143	4	US-09-301-593-26

28	64	13.6	428	3	US-09-118-319-5	Sequence 5, Appl
29	64	13.6	464	1	US-08-353-400-36	Sequence 36, Appl
30	64	13.6	472	4	US-09-301-593-30	Sequence 30, Appl
31	63.5	13.5	1785	4	US-09-341-587-3	Sequence 3, Appl
32	63	13.4	453	4	US-09-301-593-18	Sequence 18, Appl
33	63	13.4	472	4	US-09-301-593-43	Sequence 43, Appl
34	63	13.4	718	4	US-09-328-352-4640	Sequence 4640, Ap
35	63	13.4	1581	3	US-09-110-517-2	Sequence 2, Appl
36	62.5	13.3	288	4	US-09-423-439-38	Sequence 38, Appl
37	62.5	13.3	445	1	US-08-353-400-33	Sequence 33, Appl
38	62.5	13.3	641	4	US-09-687-538B-8	Sequence 8, Appl
39	62.5	13.3	673	4	US-09-423-439-32	Sequence 32, Appl
40	62.5	13.3	802	3	US-09-081-345-18	Sequence 18, Appl
41	62	13.2	1290	1	US-08-470-350B-2	Sequence 2, Appl
42	61.5	13.1	128	1	US-07-946-421-26	Sequence 26, Appl
43	61.5	13.1	307	2	US-08-484-905-68	Sequence 68, Appl
44	61.5	13.1	307	3	US-08-481-985B-68	Sequence 68, Appl
45	61.5	13.1	307	3	US-08-370-476-68	Sequence 68, Appl

ALIGNMENTS

RESULT 1

US-09-136-574A-47

; Sequence 47, Application US/09136574A

; Patent No. 6294366

; GENERAL INFORMATION:

; APPLICANT: Farrington, Graham K.

; Anderson, Paige

; Gibbs, Moreland

; Bergquist, Peter

; Daniels, Roy

; Morgan, Hugh W.

; Williams, Diane P.

; TITLE OF INVENTION: Compositions and Methods for

; Treating Cellulose Containing

; Cellulase Enzyme Compositions

; NUMBER OF SEQUENCES: 49

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Howson and Howson

; STREET: Spring House Corporate Center, P.O. Box 457

; CITY: Spring House

; STATE: PA

; COUNTRY: USA

; ZIP: 19477

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/136,574A

; FILING DATE: 19-Aug-1998

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/932,571

; FILING DATE: September 19, 1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Bak, Mary E.

; REGISTRATION NUMBER: 31,215

; REFERENCE/DOCKET NUMBER: 1997US001/CIP

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 215-540-9200

; TELEFAX: 215-540-5818

; TELEX: <Unknown>

; INFORMATION FOR SEQ ID NO: 47:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 616 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; SEQUENCE DESCRIPTION: SEQ ID NO: 47:



```

US-07-862-588B-2
; Sequence 2, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linnaa
; APPLICANT: Schlein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 59167960 No. 5916796disk of No. 5916796th America, Inc.
; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
; REFERENCE/DOCKET NUMBER: 3425.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 700 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-862-588B-2

Query Match 36.7%; Score 172; DB 2; Length 700;
Best Local Similarity 44.9%; Pred. No. 1.1e-10;
Matches 40; Conservative 18; Mismatches 27; Indels 4; Gaps 3;

QY 1 VSGGVKVOYKND-SAPGDNQIKPGLQLVNTGSSVDLSTVTVRVWFTRDGSSSTLVYNC 59
DB 549 VNSDLWQYKQDNRNATDNQIKPFNTQNGTSPVLSLTLRYFTKD-SSAAMNGWI 607
QY 60 DWAAMGCGNIRASFGSNVNPATPTADTYLQ 88
DB 608 DWAKLGSGNIQISFGNHA--DSDTYAE 634

RESULT 5
US-09-033-537A-1
; Sequence 1, Application US/09033537A
; Patent No. 5958083
; GENERAL INFORMATION:
; APPLICANT: Onishi, Masahiro
; APPLICANT: Fich, Merete
; APPLICANT: Toft, Annette Hanne
; APPLICANT: Shlein, Martin
; TITLE OF INVENTION: Prevention Of Back-Staining
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5958083o No. 5958083disk of No. 5958083th America, Inc.

```

```

; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/033,537A
; FILING DATE: 02-MAR-1998
; CLASSIFICATION: 008
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 0993/95
; FILING DATE: 08-SEP-1995
; APPLICATION NUMBER: PCT/DK96/00364
; FILING DATE: 03-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Green, Reza
; REGISTRATION NUMBER: 38,475
; REFERENCE/DOCKET NUMBER: 4492.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 551 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-033-537A-1

Query Match 35.3%; Score 165.5; DB 2; Length 551;
Best Local Similarity 37.9%; Pred. No. 4.1e-10;
Matches 33; Conservative 22; Mismatches 29; Indels 3; Gaps 2;

QY 2 SGGVKVOYKND-SAPGDNQIKPGLQLVNTGSSVDLSTVTVRVWFTRDGSSSTLVYNC 61
DB 402 TGNLVVQYKVGDTSATDNQMKPSFNKNNGTTPVNLGSLKRLRYFTKD-GTADMSASPDW 460
QY 62 AAMGCGNIRASFGSNVNPATPTADTYLQ 88
DB 461 AQIGASNVSAF--ANFTGSNTDTYVE 485

RESULT 6
PCT-US95-13813-9
; Sequence 9, Application PC/TUS9513813
; GENERAL INFORMATION:
; APPLICANT: Yeda Research and Development Co. Ltd.
; APPLICANT: Ramot University Authority for Applied
; APPLICANT: Research and Industrial Development Ltd.
; APPLICANT: Technion Research and Development Foundation Ltd.
; APPLICANT: Bayer, Edward A.
; APPLICANT: Morag, Ely
; APPLICANT: Wilchek, Meir
; APPLICANT: Lamed, Raphael
; APPLICANT: Shoham, Yuval
; TITLE OF INVENTION: MODIFIED CELLULOSE-BINDING DOMAIN (CBD)
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brodsky and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

```
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/13813
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: BAYER=3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; TELE: 248633
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 167 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
PCT-US95-13813-9

Query Match 35.2%; Score 165; DB 5; Length 167;
Best Local Similarity 36.1%; Pred. No. 1.1e-10;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKQVKNDSAPGDNQIKPGLQLVNTGSSVDLSVTYRVYWFTRDGGSSSTLYVNC 60
DB 5 VSGNLKVEFYNSPDSITNSINPQFKVTNTGSSAIDLSKLTLYRYTYVDGQDKQTFW-CD 63
QY 61 WAAM-----GCCNIRASFGSVNPATPTADTYLQ 88
DB 64 HAAIGSGSYNGITSNVKGTFFVKSSSTNNADTYLE 100

RESULT 7
US-09-339-159B-4
; Sequence 4, Application US/09339159B
; Patent No. 6566114
; GENERAL INFORMATION:
; APPLICANT: Kauppinen, Markus
; APPLICANT: Schulein, Martin
; APPLICANT: Schnorr, Kirk
; APPLICANT: Andersen, Lene
; APPLICANT: Bjornvad, Mads
; TITLE OF INVENTION: No. 6566114el Mannanases
; FILE REFERENCE: 5440.204-US
; CURRENT APPLICATION NUMBER: US/09/339,159B
; CURRENT FILING DATE: 1999-06-24
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 4
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Bacillus
US-09-339-159B-4

Query Match 35.2%; Score 165; DB 4; Length 476;
Best Local Similarity 36.1%; Pred. No. 3.9e-10;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKQVKNDSAPGDNQIKPGLQLVNTGSSVDLSVTYRVYWFTRDGGSSSTLYVNC 60
DB 314 VSGNLKVEFYNSPDSITNSINPQFKVTNTGSSAIDLSKLTLYRYTYVDGQDKQTFW-CD 372
QY 61 WAAM-----GCCNIRASFGSVNPATPTADTYLQ 88
DB 373 HAAIGSGSYNGITSNVKGTFFVKSSSTNNADTYLE 409

RESULT 8
US-09-339-159B-4
; Sequence 4, Application US/09339159B
; Patent No. 6566114
; GENERAL INFORMATION:
; APPLICANT: Kauppinen, Markus
; APPLICANT: Schulein, Martin
; APPLICANT: Schnorr, Kirk
; APPLICANT: Andersen, Lene
; APPLICANT: Bjornvad, Mads
; TITLE OF INVENTION: No. 6566114el Mannanases
; FILE REFERENCE: 5440.204-US
; CURRENT APPLICATION NUMBER: US/09/339,159B
; CURRENT FILING DATE: 1999-06-24
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 4
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Bacillus
US-09-339-159B-4

Query Match 35.2%; Score 165; DB 4; Length 476;
Best Local Similarity 36.1%; Pred. No. 3.9e-10;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKQVKNDSAPGDNQIKPGLQLVNTGSSVDLSVTYRVYWFTRDGGSSSTLYVNC 60
DB 314 VSGNLKVEFYNSPDSITNSINPQFKVTNTGSSAIDLSKLTLYRYTYVDGQDKQTFW-CD 372
QY 61 WAAM-----GCCNIRASFGSVNPATPTADTYLQ 88
DB 373 HAAIGSGSYNGITSNVKGTFFVKSSSTNNADTYLE 409

RESULT 9
US-09-198-955A-12
; Sequence 12, Application US/09198955A
; Patent No. 6187580
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Schulein, Lene E.
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. 6187580el Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/09/198,955A
; CURRENT FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 493
```

```
; TYPE: PRT
; ORGANISM: Clostridium thermocellum
US-09-198-955A-12

Query Match          35.2%; Score 165; DB 3; Length 493;
Best Local Similarity 36.1%; Pred. No. 4.1e-10;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKQYKNDSPAGDNQIKPGLQLVNTGSSVDSLSTVTVRYWFTTRDGGSSLVYNCD 60
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 331 VSGNLKVEFYNSPDSITNSINPQKVTNTGSSAIDLSKLTLYYYTVDGKQDTFW-CD 389

QY 61 WAAM-----CCGNIRASFGSVNPATPTADTYLQ 88
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
Db 390 HAAIIGSGSYNGITSNVKGTFFVKMSSSTNNADTYLE 426

RESULT 10
US-09-694-531-12
; Sequence 12, Application US/09694531
; Patent No. 6368843
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. 6368843el Pectate Lyases
; FILE REFERENCE: 5378-200-US
; CURRENT APPLICATION NUMBER: US/09/694,531
; CURRENT FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 09/198,955
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Clostridium thermocellum
US-09-694-531-12

Query Match          35.2%; Score 165; DB 4; Length 493;
Best Local Similarity 36.1%; Pred. No. 4.1e-10;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKQYKNDSPAGDNQIKPGLQLVNTGSSVDSLSTVTVRYWFTTRDGGSSLVYNCD 60
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 331 VSGNLKVEFYNSPDSITNSINPQKVTNTGSSAIDLSKLTLYYYTVDGKQDTFW-CD 389

QY 61 WAAM-----CCGNIRASFGSVNPATPTADTYLQ 88
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 390 HAAIIGSGSYNGITSNVKGTFFVKMSSSTNNADTYLE 426

RESULT 11
US-09-670-141-10
; Sequence 10, Application US/09670141
```

```
; Patent No. 6429000
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels Erik K.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Schnorr, Kirk
; TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus
; FILE REFERENCE: 5377-200-US
; CURRENT APPLICATION NUMBER: US/09/670,141
; CURRENT FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: 09/198,956
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Bacillus licheniformis
US-09-670-141-10

Query Match          35.2%; Score 165; DB 4; Length 493;
Best Local Similarity 36.1%; Pred. No. 4.1e-10;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKQYKNDSPAGDNQIKPGLQLVNTGSSVDSLSTVTVRYWFTTRDGGSSLVYNCD 60
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 331 VSGNLKVEFYNSPDSITNSINPQKVTNTGSSAIDLSKLTLYYYTVDGKQDTFW-CD 389

QY 61 WAAM-----CCGNIRASFGSVNPATPTADTYLQ 88
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 390 HAAIIGSGSYNGITSNVKGTFFVKMSSSTNNADTYLE 426

RESULT 12
US-10-072-152-12
; Sequence 12, Application US/10072152
; Patent No. 6677147
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. 6677147el Pectate Lyases
; FILE REFERENCE: 5378-200-US
; CURRENT APPLICATION NUMBER: US/10/072,152
; CURRENT FILING DATE: 2002-02-07
; PRIOR APPLICATION NUMBER: US/09/198,955
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 12
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Clostridium thermocellum
US-10-072-152-12

Query Match      35.2%; Score 165; DB 4; Length 493;
Best Local Similarity 36.1%; Pred. No. 4.1e-10;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKQYKNDSPGDNQIKPGLQLVNTGSSVDLSTVTYVYFTRDGGSSSTLVYNC 60
Db 331 VSGNLKVEFYNSPDDTNSINPQKVTNGSSADLSKLTURLYTYTVDGQKQTFW-CD 389

QY 61 WAAM-----GCGNIRASFGSVNPATPTADTYLQ 88
Db 390 HAAIIGNSYNGITSNVKGFVKSSSTNNADTYLE 426

RESULT 13
US-07-862-588B-7
; Sequence 7, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linea
; APPLICANT: Sch lein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESS: No. 59167960 No. 5916796disk of No. 5916796th America, Inc.
; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
; REFERENCE/DOCKET NUMBER: 3425.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 531 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-862-588B-7

Query Match      25.7%; Score 120.5; DB 2; Length 531;
Best Local Similarity 32.9%; Pred. No. 3.5e-05;
Matches 28; Conservative 18; Mismatches 36; Indels 3; Gaps 2;

QY 2 SGGVKVQYKNDSPGDNQIKPGLQLVNTGSSVDLSTVTYVYFTRDGGSSSTLVYNC 61
Db 432 TGNLVVQYKVGDTSATDNQKPSFNKNGTTPVNLGSLKXXXXKD-GPADMSCSIDW 490
```

```
QY 62 AAMGCGNIRASFGSVNPATPTADTY 86
Db 491 AQIGRTNVLAP--ANFTGSNTDTY 513

RESULT 14
US-08-048-164A-2
; Sequence 2, Application US/08048164A
; Patent No. 5496934
; GENERAL INFORMATION:
; APPLICANT: Shoseyov, Oded
; APPLICANT: Shpiegl, Itai
; APPLICANT: Goldstein, Marc A.
; APPLICANT: Doi, Roy H.
; TITLE OF INVENTION: CELLULOSE BINDING DOMAIN
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESS: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/048,164A
; FILING DATE: 14-APR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7809-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 162 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-048-164A-2

Query Match      24.3%; Score 114; DB 1; Length 162;
Best Local Similarity 29.9%; Pred. No. 4.2e-05;
Matches 29; Conservative 22; Mismatches 34; Indels 12; Gaps 4;

QY 2 SGGVKVQYKNDSPGDNQIKPGLQLVNTGSSVDLSTVTYVYFTRDGGSSSTLVYNC 61
Db 3 TSSMSVEFYNSKSAQTSITPIKITNTSDNLNLDVKRYVYTSDTGQTGFW-CDH 61

QY 62 AAMGCGN-----IRASF--GSVNPATPTADTYLQ 88
Db 62 AGALLGNSYVDNTSKVTANFVKETASP-TSTYDTYVE 97

RESULT 15
US-08-460-462-2
; Sequence 2, Application US/08460462
; Patent No. 5670623
; GENERAL INFORMATION:
; APPLICANT: Shoseyov, Oded
; APPLICANT: Shpiegl, Itai
; APPLICANT: Goldstein, Marc A.
; APPLICANT: Doi, Roy H.
; TITLE OF INVENTION: METHODS OF USE OF CELLULOSE BINDING DOMAIN PROTEINS
; NUMBER OF SEQUENCES: 21
```

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,462
FILING DATE: concurrently herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/048,164
FILING DATE: 14-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Misrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7809-006
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 162 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-460-462-2

```

```

Query Match      24.3%; Score 114; DB 1; Length 162;
Best Local Similarity 29.9%; Pred. NO. 4.2e-05;
Matches 29; Conservative 22; Mismatches 34; Indels 12; Gaps 4;

Qy      2 SGGVKVQYKNDSPAGDNOIKPGQLVNTGSSVDLSTVTYVWFTRDGSGSTLYVNCDW 61
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
          3 TSMSVFEFNSKSAQNTSITPIIKITNTSDSLNLDNDVKRYYYTSDGTQGTFW-CDH 61
Qy      62 AAMGCGN-----IRASF--GSVNPATPTADTVIQ 88
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
          62 AGALLGNSYDNTSKVTANFVKETASP-TSTYDTYVE 97

```

Search completed: May 11, 2004, 12:11:22  
Job time : 5.31843 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 12:09:36 ; Search time 10.2683 Seconds  
(without alignments)  
2378.773 Million cell updates/sec

Title: US-09-917-376-5

Perfect score: 459

Sequence: 1 VSGGVKVKYKNDSPAGDNQ.....IRASFGSVNPATPTADTYLQ 88

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1140673 seqs, 277566755 residues

Total number of hits satisfying chosen parameters: 1140673

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Published Applications AA:\*
- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
  - 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
  - 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
  - 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
  - 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
  - 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
  - 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
  - 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
  - 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*
  - 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
  - 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
  - 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
  - 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
  - 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
  - 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
  - 16: /cgn2\_6/ptodata/2/pubpaa/US10C\_NEW\_PUB.pep.\*
  - 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
  - 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	469	100.0	88	12	US-09-917-376-5
2	469	100.0	88	14	US-10-155-400-5
3	469	100.0	89	12	US-09-917-376-4
4	469	100.0	89	14	US-10-155-400-4
5	469	100.0	154	10	US-09-917-378-4
6	469	100.0	762	10	US-09-917-378-1
7	466	99.4	150	10	US-09-917-384-5
8	466	99.4	150	10	US-09-917-383-5
9	466	99.4	1043	10	US-09-917-384-6
10	466	99.4	1043	10	US-09-917-383-6
11	466	99.4	1228	10	US-09-917-384-1
12	466	99.4	1228	10	US-09-917-383-1
13	462	98.5	957	12	US-09-917-376-1
14	462	98.5	957	14	US-10-155-400-1
15	257	54.8	741	14	US-10-156-761-8100

16	190	40.5	508	15	US-10-369-493-23151	Sequence 23151, A
17	180	38.4	1621	14	US-10-185-990-10	Sequence 10, Appl
18	165	35.2	476	12	US-10-372-054-4	Sequence 4, Appl
19	165	35.2	493	13	US-10-655-433-12	Sequence 12, Appl
20	165	35.2	493	13	US-10-072-152-12	Sequence 12, Appl
21	165	35.2	599	10	US-09-955-555A-29	Sequence 29, Appl
22	155.5	33.2	1352	10	US-09-784-554B-2	Sequence 2, Appl
23	150.5	32.1	1350	10	US-09-784-554B-4	Sequence 4, Appl
24	130.5	27.8	1483	12	US-10-282-122A-51483	Sequence 51483, A
25	114	24.3	256	14	US-10-261-446-6	Sequence 22, Appl
26	113	24.1	382	10	US-09-808-898-22	Sequence 2, Appl
27	105.5	22.5	163	12	US-10-460-524-2	Sequence 2, Appl
28	78.5	16.7	1049	12	US-10-282-122A-49900	Sequence 49900, A
29	75.5	16.1	618	12	US-10-211-462-223	Sequence 23, Appl
30	75.5	16.1	986	9	US-09-747-835A-53	Sequence 53, Appl
31	75.5	16.1	986	12	US-10-312-312-53	Sequence 53, Appl
32	75.5	16.1	986	14	US-10-120-604-101	Sequence 101, Appl
33	75.5	16.1	986	14	US-10-225-567A-406	Sequence 406, Appl
34	75.5	16.1	986	16	US-10-398-458-3	Sequence 3, Appl
35	75.5	16.1	1131	16	US-10-398-458-2	Sequence 2, Appl
36	75.5	16.1	1325	9	US-09-747-835A-24	Sequence 24, Appl
37	75.5	16.1	1325	12	US-10-312-312-24	Sequence 24, Appl
38	75.5	16.1	1346	9	US-09-747-835A-20	Sequence 20, Appl
39	75.5	16.1	1346	9	US-09-747-835A-61	Sequence 61, Appl
40	75.5	16.1	1346	12	US-10-312-312-20	Sequence 20, Appl
41	75.5	16.1	1346	12	US-10-312-312-61	Sequence 61, Appl
42	75.5	16.1	1371	9	US-09-747-835A-25	Sequence 25, Appl
43	75.5	16.1	1371	12	US-10-312-312-25	Sequence 25, Appl
44	75	16.0	161	12	US-10-424-599-136885	Sequence 196885, A
45	72.5	15.5	69	9	US-09-764-860-377	Sequence 377, Appl

ALIGNMENTS

RESULT 1

US-09-917-376-5  
; Sequence 5, Application US/09917376  
; Publication No. US20040038334A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOLYTICUS  
; FILE REFERENCE: 40197.4US01  
; CURRENT APPLICATION NUMBER: US/09/917,376  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 88  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
; FEATURE:  
; OTHER INFORMATION: Carbohydrate binding domain

Query Match 100.0%; Score 469; DB 12; Length 88;  
Best Local Similarity 100.0%; Pred. No. 1.5e-48;  
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKVKYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTYVWFTRDGGSTLVNCD 60  
Db 1 VSGGVKVKYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTYVWFTRDGGSTLVNCD 60

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
Db 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88

RESULT 2

```
US-10-155-400-5
; Sequence 5, Application US/10155400
; Publication No. US2003010898A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: NREL 01-36A
; CURRENT APPLICATION NUMBER: US/10/155,400
; CURRENT FILING DATE: 2002-10-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 88
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; OTHER INFORMATION: Carbohydrate binding domain
US-10-155-400-5

Query Match      100.0%; Score 469; DB 14; Length 88;
Best Local Similarity 100.0%; Pred. No. 1.5e-48;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQYKKNDSAPGDNQIKPGLQVNTGSSVDLSVTYVYFTRDGGSSSTLYNCD 60
Db 1 VSGGVKQYKKNDSAPGDNQIKPGLQVNTGSSVDLSVTYVYFTRDGGSSSTLYNCD 60

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
Db 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88

RESULT 3
US-09-917-376-4
; Sequence 4, Application US/09917376
; Publication No. US2004003833A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: 40197.4US01
; CURRENT APPLICATION NUMBER: US/09/917,376
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 89
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (89)
; OTHER INFORMATION: Any amino acid
; OTHER INFORMATION: Carbohydrate binding domain
US-09-917-376-4

Query Match      100.0%; Score 469; DB 12; Length 89;
Best Local Similarity 100.0%; Pred. No. 1.5e-48;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQYKKNDSAPGDNQIKPGLQVNTGSSVDLSVTYVYFTRDGGSSSTLYNCD 60
Db 1 VSGGVKQYKKNDSAPGDNQIKPGLQVNTGSSVDLSVTYVYFTRDGGSSSTLYNCD 60

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
Db 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88

US-10-155-400-4
; Sequence 4, Application US/10155400
; Publication No. US2003010898A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: NREL 01-36A
; CURRENT APPLICATION NUMBER: US/10/155,400
; CURRENT FILING DATE: 2002-10-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 89
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (89)
; OTHER INFORMATION: Any amino acid
; OTHER INFORMATION: Carbohydrate binding domain
US-10-155-400-4

Query Match      100.0%; Score 469; DB 14; Length 89;
Best Local Similarity 100.0%; Pred. No. 1.5e-48;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQYKKNDSAPGDNQIKPGLQVNTGSSVDLSVTYVYFTRDGGSSSTLYNCD 60
Db 1 VSGGVKQYKKNDSAPGDNQIKPGLQVNTGSSVDLSVTYVYFTRDGGSSSTLYNCD 60

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
Db 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88

RESULT 5
US-09-917-378-4
; Sequence 4, Application US/09917378
; Publication No. US20030119093A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT MANNANASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: 40197.7US01
; CURRENT APPLICATION NUMBER: US/09/917,378
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 154
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Carbohydrate
; OTHER INFORMATION: binding domain
US-09-917-378-4

Query Match      100.0%; Score 469; DB 10; Length 154;
Best Local Similarity 100.0%; Pred. No. 2.9e-48;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy	1	VSGGVQVQYKNDSAPGDNQI1KPGQLVNTGSSVDLSTVTVRYNFTRDGSSSTLVYNCD	60
Db	1	VSGGVQVQYKNDSAPGDNQI1KPGQLVNTGSSVDLSTVTVRYNFTRDGSSSTLVYNCD	60
Qy	61	WAAMGCNIRASFGSVNPATPTADTYLQ	88
Db	61	WAAMGCNIRASFGSVNPATPTADTYLQ	88

```

RESULT 6
US-09-917-378-1
; Sequence 1, Application US/09917378
; Publication No. US20030119093A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT MANNANASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: 40197.7US01
; CURRENT APPLICATION NUMBER: US/09/917,378
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 762
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
US-09-917-378-1

```

```

Query Match      100.0%; Score 469; DB 10; Length 762;
Best Local Similarity 100.0%; Pred. No. 2e-47;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1 VSGGVKVQYKNDSPAGDNIQIKPGLQLVNTGSSSVDLSTVTRYWFTRDGGSSTLVNCD 60
      |||||
ddb     455 VSGGVKVQYKNDSPAGDNIQIKPGLQLVNTGSSSVDLSTVTRYWFTRDGGSSTLVNCD 514
      |||||

Oy      61 WAAMGCCNIRASFGSVNPATPDTYIQ 88
      |||||
ddb     515 WAAMGCCNIRASFGSVNPATPDTYIQ 542
      |||||

```

RESULT 7  
 US-09-917-384-5  
 ; Sequence 5, Application US/09917384  
 ; Publication No. US20030096342A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: DING, SHI-YOU  
 ; APPLICANT: ADNEY, WILLIAM S.  
 ; APPLICANT: VINZANT, TODD B.  
 ; APPLICANT: DECKER, STEPHEN R.  
 ; APPLICANT: HIMMEL, MICHAEL E.  
 ; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS  
 ; TITLE OF INVENTION: CELLULOLYTICUS  
 ; FILE REFERENCE: 40170.6US01  
 ; CURRENT APPLICATION NUMBER: US/09/917,384  
 ; CURRENT FILING DATE: 2001-07-28  
 ; NUMBER OF SEQ ID NOS: 14  
 ; SOFTWARE: PatentIn ver. 2.1  
 ; SEQ ID NO 5  
 ; LENGTH: 150  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: Segment C  
 ; OTHER INFORMATION: Guxa  
 ; US-09-917-384-5

Query Match 99.4%; Score 466; DB 10; Length 150;

	Best Local Similarity	98.9%;	Pred. No. 6.5e-48;			
	Matches	87;	Conservative	1;	Mismatches	0;
				Indels	0;	Gaps
Qy	1	VSGGVKQVKYKNNDSAPGDNQIKPGLQIVNTGSSVDLSTVTVRYWFTRDGGSSTLVNCD	60			
Db	1	VSGGLKQVKYKNNDSAPGDNQIKPGLQIVNTGSSVDLSTVTVRYWFTRDGGSSTLVNCD	60			
Qy	61	WAAMGCCNTRASFGSVNPTPTADTYLQ	88			
Db	61	WAAMGCCNTRASFGSVNPTPTADTYLO	88			

```

RESULT 8
US-09-917-383-5
; Sequence 5, Application US/09917383
; Publication No. US20030104522A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOXYLICUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917.383
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 5
; LENGTH: 150
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
US-09-917-383-5

```

```

Query Match      99.4%; Score 466; DB 10; Length 150;
Best Local Similarity 98.9%; Pred. No. 6.5e-48;
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 VSGGVKVQYKNDSPAGDNDQIKPGLQVNTGSSVDLSTTVYRYWFFTRDGSSTLVNCD 60
      |||:|||||
Db      1 VSGGLKVQYKNDSPAGDNDQIKPGLQVNTGSSVDLSTTVYRYWFFTRDGSSTLVNCD 60
      |||:|||||

QY      61 WAAMCGGNIRASFGSVNPATPTADTYLQ 88
      |||:|||||
Db      61 WAAMCGGNIRASFGSVNPATPTADTYLQ 88
      |||:|||||

```

RESULT 9  
US-09-917-384-6  
; Sequence 6, Application US/09917384  
; Publication No. US20030096342A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: DECKER, STEPHEN R.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOLYTICUS  
; FILE REFERENCE: 40170.6US01  
; CURRENT APPLICATION NUMBER: US/09/917,384  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 1043  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:



; CURRENT APPLICATION NUMBER: US/09/917,376  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 957  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
; FEATURE:  
; NAME/KEY: MOD RES  
; LOCATION: (957)  
; OTHER INFORMATION: Any amino acid  
US-09-917-376-1

Query Match 98.5%; Score 462; DB 12; Length 957;  
Best Local Similarity 97.7%; Pred. No. 1.9e-46;  
Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VSGGVKQYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSSTLVYNC 60  
DB 869 VSGGVKQYKNDAPGDNQIKPGLQVVNTGSSVDLSTVTYRYWFTRDGSSSTLVYNC 928  
QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
DB 929 WAAIGCGNIRASFGSVNPATPTADTYLQ 956

RESULT 14  
US-10-155-400-1  
; Sequence 1, Application US/10155400  
; Publication No. US20030108988A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOXYLICUS  
; FILE REFERENCE: NREL 01-36A  
; CURRENT APPLICATION NUMBER: US/10/155,400  
; CURRENT FILING DATE: 2002-10-22  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 957  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
; FEATURE:  
; NAME/KEY: MOD RES  
; LOCATION: (957)  
; OTHER INFORMATION: Any amino acid  
US-10-155-400-1

Query Match 98.5%; Score 462; DB 14; Length 957;  
Best Local Similarity 97.7%; Pred. No. 1.9e-46;  
Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VSGGVKQYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSSTLVYNC 60  
DB 869 VSGGVKQYKNDAPGDNQIKPGLQVVNTGSSVDLSTVTYRYWFTRDGSSSTLVYNC 928  
QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
DB 929 WAAIGCGNIRASFGSVNPATPTADTYLQ 956

RESULT 15  
US-10-156-761-8100  
; Sequence 8100, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO

; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 8100  
; LENGTH: 741  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-8100  
Query Match 54.8%; Score 257; DB 14; Length 741;  
Best Local Similarity 56.3%; Pred. No. 6e-22;  
Matches 49; Conservative 13; Mismatches 25; Indels 0; Gaps 0;  
QY 2 SGGVKVQYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSSTLVYNC 61  
DB 591 SGGKLVLYKNDSSATDNARPGLRIVNTGSGSLDLKVTARYYFTRDGSGSPTVNAWCDY 650  
QY 62 AAMCGNIRASFGSVNPATPTADTYLQ 88  
DB 651 AAVGCSNVSLKVVPLTTPVPGADAYLE 677

Search completed: May 11, 2004, 12:25:48  
Job time : 10.2883 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 12, 2004, 03:40:05 ; Search time 1264.92 Seconds  
(without alignments)  
3015.368 Million cell updates/sec

Title: US-09-917-376-5  
Perfect score: 469  
Sequence: 1 VSGGVKQVYKNDSAPGDNQ.....IRASFGSVNPATPTADTYLQ 88

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame\_plus\_p2n.model -DEV=xlh  
-Q=/Cgn2\_1/USPIO\_gpool/US09917376/runat\_11052004\_114535\_28241/app\_query.fasta\_1.1429  
-DB=GenEmbl -QWTF=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45  
-LOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09917376 @CGN 1 1 7742 @runat\_11052004\_114535\_28241 -NCPU=6 -ICPU=3  
-NO WMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*

- 29: em.vi.\*
- 30: em.htg\_hum.\*
- 31: em.htg\_inv.\*
- 32: em.htg\_other.\*
- 33: em.htg\_mus.\*
- 34: em.htg\_pln.\*
- 35: em.htg\_rod.\*
- 36: em.htg\_mam.\*
- 37: em.htg\_vrt.\*
- 38: em.sv.\*
- 39: em.htgo\_hum.\*
- 40: em.htgo\_mus.\*
- 41: em.htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	469	100.0	2289	6	AX700050	Sequence
2	466	99.4	3687	6	AX700036	Sequence
3	462	98.5	2869	6	AX700058	Sequence
4	436	93.0	3365	6	AX700025	Sequence
5	257	54.8	299175	1	AP005023	Streptomy
6	228.5	48.7	1957	1	AY298814	Thermobif
7	225.5	48.1	4567	1	AF163837	Caldibaci
8	214.5	45.7	3237	1	AF200304	Caldibaci
9	209.5	44.7	3509	1	PAE488933	Paenibaci
10	204.5	43.6	4161	1	BSP133614	Bacillus
11	195.5	41.7	2029	6	E35142	Truncated c
12	195.5	41.7	2029	6	E35143	Truncated c
13	194.5	41.5	6005	1	AF078038S5	Caldicell
14	194.5	41.5	6416	6	E35100	Truncated c
15	192.5	41.0	4743	1	AB008029	Thermophi
16	192.5	41.0	11707	1	AF078737	Caldicell
17	192.5	41.0	11707	6	E35099	Truncated c
18	191.5	40.8	787	1	AF078038S3	Caldicell
19	191	40.7	5513	1	ATZ86105	A.thermophi
20	190.5	40.6	3262	1	AF078038S1	Caldicell
21	190.5	40.6	5437	1	CASR69XYN2	Caldicell
22	190	40.5	1523	1	AF355629	Bacillus
23	190	40.5	1800	1	BSEGLUG	Bacillus su
24	190	40.5	1920	1	BACGLUB	B. subtilis
25	190	40.5	2314	1	BS14GLUC	B. subtilis
26	190	40.5	2435	1	BSBGLUC2	Bacillus su
27	190	40.5	2589	1	AY044252	Bacillus
28	190	40.5	26170	1	BC170DEGR	B. subtilis
29	190	40.5	207829	1	BSUB0010	Bacillus su
30	188	40.1	1928	1	BACMCASE	B. subtilis
31	187	39.9	1593	1	AY183475	Bacillus
32	187	39.9	2084	1	AF045482	Bacillus
33	186.5	39.8	4977	1	CDCWANA	Caldocellum
34	186.5	39.8	5439	1	CDCCELA	Caldocellum
35	186	39.7	2175	6	BACCELD	B. subtilis
36	186	39.7	2175	6	E05425	DNA sequenc
37	186	39.7	2587	1	AB016164	Bacillus
38	182.5	38.9	2977	6	A28170	B. lautus st
39	182.5	38.9	4241	1	CSCELB	Caldocellum
40	182.5	38.9	5284	1	CSU16308	Caldocellum
41	180.5	38.5	2831	1	BACCELAA	Bacillus la
42	180	38.4	1553	1	BSU27084	Bacillus sp
43	177	37.7	2152	1	AF363635	Bacillus
44	172.5	36.8	6555	1	AF155197	Acetivibr
45	168	35.8	291000	1	SCO939105	Streptomy

ALIGNMENTS

RESULT 1

```

AX700050
LOCUS AX700050 2289 bp DNA linear PAT 03-APR-2003
DEFINITION Sequence 2 from Patent WO03012110.
ACCESSION AX700050
VERSION AX700050.1 GI:29536020
KEYWORDS
SOURCE
ORGANISM
Acidothermus cellulolyticus
Acidothermus cellulolyticus
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Frankineae; Acidothermaceae; Acidothermus.
REFERENCE
1
AUTHORS Ding, S.Y., Adney, W.S., Vinzant, T.B. and Himmel, M.E.
TITLE Thermal tolerant mannanase from Acidothermus cellulolyticus
JOURNAL Patent: WO 03012109-A 2 13-FEB-2003;
Midwest Research Institute (US)
FEATURES
source
1..2289
/organism="Acidothermus cellulolyticus"
/mol_type="unassigned DNA"
/db_xref="taxon:28049"
ORIGIN
Alignment Scores:
Pred. No.: 2,21e-48 Length: 2289
Score: 469.00 Matches: 88
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
DB: 6
US-09-917-376-5 (1-88) x AX700050 (1-2289)
QY 1 ValSerGlyGlyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 1363 GTGTCGGGTGGGTGAAGTGCAGTACAGAACAAATGATTCGGCGCGGGTGATACCAG 1422
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1423 ATCAAAACCGGGTCTCCAGTTGGTGAATACGGGGTCGTCGGTGGATTGTCGACGGTG 1482
QY 41 ThrValArgTyrTrpPheThrArgAspGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1483 ACGGTGCGGTACTGGTTCACCGGGATGGTGGTCTGTCGACACTGGTGTACAACTGTGAC 1542
QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 1543 TGGCGCGCGCATGGGTGTGGGAATATCCGGGCTCGTTCGGTCCGTTGAACCCGGCGACG 1602
QY 81 ProThrAlaAspThrTyrLeuGln 88
Db 1603 CCGACGGCGGACACCTACCTGCAG 1626
RESULT 2
AX700036
LOCUS AX700036 3687 bp DNA linear PAT 03-APR-2003
DEFINITION Sequence 2 from Patent WO03012109.
ACCESSION AX700036
VERSION AX700036.1 GI:29536019
KEYWORDS
SOURCE
ORGANISM
synthetic construct
synthetic construct
artificial sequences.
REFERENCE
1
AUTHORS Ding, S.Y., Adney, W.S., Vinzant, T.B., Himmel, M.E. and Decker, S.R.
TITLE Thermal tolerant cellulase from Acidothermus cellulolyticus
JOURNAL Patent: WO 03012109-A 2 13-FEB-2003;
Midwest Research Institute (US)
FEATURES
source
1..3687
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="Segment of GuxA"

```

```

ORIGIN
Alignment Scores:
Pred. No.: 9.53e-48 Length: 3687
Score: 466.00 Matches: 87
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.86% Mismatches: 0
Query Match: 99.36% Indels: 0
Gaps: 0
DB: 6
US-09-917-376-5 (1-88) x AX700036 (1-3687)
QY 1 ValSerGlyGlyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 1750 GTGTCGGGTGGGTGAAGTGCAGTACAGAACAAATGATTCGGCGCGGGTGATAACCAG 1809
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1810 ATCAAAACCGGGTCTCCAGTTGGTGAATACCGGGTCGTCGGTGGATTGTCGACGGTG 1869
QY 41 ThrValArgTyrTrpPheThrArgAspGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1870 ACGGTGCGGTACTGGTTCACCGGGATGGTGGTCTGTCGACACTGGTGTACAACTGTGAC 1929
QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 1930 TGGCGCGCGCATGGGTGTGGGAATATCCGGGCTCGTTCGGGCTCGTGAACCCGGCGACG 1989
QY 81 ProThrAlaAspThrTyrLeuGln 88
Db 1990 CCGACGGCGGACACCTACCTGCAG 2013
RESULT 3
AX700058
LOCUS AX700058 2869 bp DNA linear PAT 03-APR-2003
DEFINITION Sequence 2 from Patent WO03012090.
ACCESSION AX700058
VERSION AX700058.1 GI:29536021
KEYWORDS
SOURCE
ORGANISM
Acidothermus cellulolyticus
Acidothermus cellulolyticus
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Frankineae; Acidothermaceae; Acidothermus.
REFERENCE
1
AUTHORS Ding, S.Y., Adney, W.S., Vinzant, T.B. and Himmel, M.E.
TITLE Thermal tolerant avicelase from Acidothermus cellulolyticus
JOURNAL Patent: WO 03012090-A 2 13-FEB-2003;
Midwest Research Institute (US)
FEATURES
source
1..2869
/organism="Acidothermus cellulolyticus"
/mol_type="unassigned DNA"
/db_xref="taxon:28049"
ORIGIN
Alignment Scores:
Pred. No.: 2.19e-47 Length: 2869
Score: 462.00 Matches: 86
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 97.73% Mismatches: 0
Query Match: 98.51% Indels: 0
Gaps: 0
DB: 6
US-09-917-376-5 (1-88) x AX700058 (1-2869)
QY 1 ValSerGlyGlyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 2605 GTGTCGGGTGGGTGAAGTGCAGTACAGAACAAATGATTCGGCGCGGGTGATAATCAG 2664
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 2665 ATCAAGCCGGGTTTTCGAGGTGGTGAATACCGGGTCGTCGGTGGATTGTCGACGGTG 2724

```

```

QY      41  ThrValArgTyrTrrPheThrArgAspGlyGlySerSerThrLeuValTyrIasnCysAsp 60
Db      2725  ACCGGTCGGTACTGTTTACC CGGGATGCTGCTCGACACTGCTGTACAACTGTGAC 2784

QY      61  TrrAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValIasnProAlaThr 80
Db      2785  TGGGGCGGATCGGGTGTGGGAATATCCGCGCTCTGTTCCGTCGGTGAACCGCGGACG 2844

QY      81  ProThrAlaAspThrTyrLeuGln 88
Db      2845  CCGACGGCGACACTACTCTGCAG 2868

RESULT 4
AX700025
LOCUS      AX700025          3365 bp      DNA      linear      PAT 03-APR-2003
DEFINITION Sequence 2 from Patent WO03012095.
ACCESSION  AX700025
VERSION     AX700025.1  GI:29536018
KEYWORDS
SOURCE
ORGANISM   Acidothermus cellulolyticus
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Frankineae; Acidothermaceae; Acidothermus.
REFERENCE
AUTHORS    Adney,W.S., Ding,S.Y., Vinzant,T.B., Himmel,M.E., Decker,S.R. and
            Lantz McCarter,S.
TITLE      Thermal tolerant exoglucanase from Acidothermus cellulolyticus
JOURNAL    Patent: WO 03012095-A 2 13-FEB-2003;
            Midwest Research Institute (US)
FEATURES   source
            1..3365
            /organism="Acidothermus cellulolyticus"
            /mol_type="unassigned DNA"
            /db_xref="taxon:28049"

ORIGIN
Alignment Scores:
Pred. No.:      4, 72e-44      Length:      3365
Score:          436.00      Matches:      81
Percent Similarity: 97.62%      Conservative: 1
Best local Similarity: 96.43%      Mismatches: 2
Query Match:     92.96%      Indels:      0
DB:              6      Gaps:      0

US-09-917-376-5 (1-88) x AX700025 (1-3365)

QY      5  VallysValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGlnIleLysProGly 24
Db      112  CTCAAAGCGCAGTATAAGAACAAATGATTCGGCGCCGAGTGACACAGATCAACCGGGT 171

QY      25  LeuGlnLeuValAsnThrGlySerSerSerValAspLeuSerThrValThrValArgTyr 44
Db      172  CTCACAGTTGGTGAATPACCGGGTCGTCGTCGGTGGATTTGTCGACGTCGCGGTAC 231

QY      45  TrrPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaIaMet 64
Db      232  TGGTTACCGCGGATGGTGGTGGTCGTCGACACTGGGTGTACACTGTGACTGGCGCGGATG 291

QY      65  GlyCysGlyAsnIleArgAlaSerPheGlySerValIasnProAlaThrProThrAlaAsp 84
Db      292  GGGTGGGAATATCCGCGCTCTGTTCCGCTCGGTGAACCCGCGGACCGCGGACG 351

QY      85  ThrTyrLeuGln 88
Db      352  ACCTACTCTGCAG 363

RESULT 5
AP005023
LOCUS      AP005023          299175 bp      DNA      linear      BCT 10-MAY-2003
DEFINITION Streptomyces avermitilis genomic DNA, complete genome, section
            3/30.
ACCESSION  AP005023 BA000030
VERSION     AP005023.1  GI:29604083

```

```

KEYWORDS
SOURCE
ORGANISM   Streptomyces avermitilis MA-4680
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE
AUTHORS    Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Hanamoto,A., Takahashi,C.,
            Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osonoe,T.,
            Kikuchi,H., Shiba,T., Sakaki,Y. and Hattori,M.
TITLE      Genome sequence of an industrial microorganism Streptomyces
            avermitilis: deducing the ability of producing secondary
            metabolites
JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
MEDLINE    21477403
PUBMED     11572948
REFERENCE
AUTHORS    Ikeda,H., Ishikawa,J., Hanamoto,A., Shinose,M., Kikuchi,H.,
            Shiba,T., Sakaki,Y., Hattori,M. and Omura,S.
TITLE      Complete genome sequence and comparative analysis of the industrial
            microorganism Streptomyces avermitilis
JOURNAL    Nat. Biotechnol. 21 (5), 526-531 (2003)
MEDLINE    22608306
PUBMED     12692562
REFERENCE
AUTHORS    Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C.,
            Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osonoe,T.,
            Kushida,N., Director-General of Biotechnology Center, Shiba,T.,
            Sakaki,Y. and Hattori,M.
TITLE      Direct Submission
JOURNAL    Submitted (29-MAR-2002) Director-General of Biotechnology Center,
            National Institute of Technology and Evaluation, Biotechnology
            Center: 2-49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan
            (E-mail:bio@nite.go.jp, URL:http://www.bio.nite.go.jp/,
            Tel:81-3-3481-1933, Fax:81-3-3481-8424)
COMMENT    This work was done in collaboration with Haruo Ikeda(*1), Jun
            Ishikawa(*2), Akiharu Hanamoto(*3), Chigusa Takahashi(*3), Mayumi
            Shinose(*3), Hiroshi Horikawa(*4), Hidekazu Nakazawa(*4), Tomomi
            Osonoe(*4), Norihiro Kushida(*4), Hisashi Kikuchi(*4), Tadayoshi
            Shiba(*5), Yoshiyuki Sakaki(*6,*7), Masahira Hattori(*1,*7)
            and Satoshi Omura(*1,*3).
            Final finishing process and all annotation were done by H. Ikeda
            and J. Ishikawa.
            *1 Kitasato Institute for Life Sciences, Kitasato University
            *2 National Institute of Infectious Diseases
            *3 The Kitasato Institute
            *4 National Institute of Technology and Evaluation
            *5 School of Science, Kitasato University
            *6 Institute of Medical Science, University of Tokyo
            *7 RIKEN, Genomic Sciences Center
            Following url is also available.
            http://avermitilis.ls.kitasato-u.ac.jp.
            Location/Qualifiers
            1..299175
            /organism="Streptomyces avermitilis MA-4680"
            /mol_type="genomic DNA"
            /db_xref="taxon:227882"
            /note="This strain is also named as strain: ATCC 31267,
            NCIMB 12804 or NRRL 8165.--synonym: Streptomyces
            avermectinius"
            complement(155..550)
            /note="SAV445
            Tn4561, truncated"
            /codon_start=1
            /transl_table=11
            /product="putative transposase"
            /protein_id="BAC68155.1"
            /db_xref="GI:29604084"
            /translation="MAGPPRSATRSRTTADREDSAHPPADRPBEGYRQIKVQANLQ
            EGHALLARAGQLYQRYQDGMEDQIGALGVNLVLFNTRYMDAALTQLRADGPEVR
            DEVARLSPFMRHHINVLGRYSFQLPDLE"
            complement(474..740)
            /note="IS630 family, truncated"
            CDS

```

```

FEATURES
source
            1..299175
            /organism="Streptomyces avermitilis MA-4680"
            /mol_type="genomic DNA"
            /db_xref="taxon:227882"
            /note="This strain is also named as strain: ATCC 31267,
            NCIMB 12804 or NRRL 8165.--synonym: Streptomyces
            avermectinius"
            complement(155..550)
            /note="SAV445
            Tn4561, truncated"
            /codon_start=1
            /transl_table=11
            /product="putative transposase"
            /protein_id="BAC68155.1"
            /db_xref="GI:29604084"
            /translation="MAGPPRSATRSRTTADREDSAHPPADRPBEGYRQIKVQANLQ
            EGHALLARAGQLYQRYQDGMEDQIGALGVNLVLFNTRYMDAALTQLRADGPEVR
            DEVARLSPFMRHHINVLGRYSFQLPDLE"
            complement(474..740)
            /note="IS630 family, truncated"
            CDS

```



```
SAV446"
/codon_start=1
/transl_table=11
/product="putative transposase"
/protein_id="BAC68156.1"
/db_xref="GI:29604085"
/translation="MPDIONAGPSTSVTAAAVSTAGSVYMMAPTSFVMAAPPSPGRRR
EFGSLGAVRAYDVIRMLSDGRPTPLGDAITYGGSRLCTSCA"
766..987
/note=".987"
/codon_start=1
/transl_table=11
/product="putative transposase"
/protein_id="BAC68157.1"
/db_xref="GI:29604086"
/translation="WVVAVGALTSVPTAGGVAAEPVCVRRLTDGQKLIQIVRRG
SASSVFRFRMMLLASAGNRVPVIAQLV"
complement(1954..2571)
/note="SAV448"
/codon_start=1
/transl_table=11
/product="putative TetR-family transcriptional regulator"
/protein_id="BAC68158.1"
/db_xref="GI:29604087"
/translation="MPTSRARERANTRERIEAALHVLTEGIAALTRIRIATDVE
YSAPVYQHFANKDALVLEVAHGRMLSEFSQAQEPDTRRMTRIASQVVRFAGE
HPHLQVMDPMVDADERVAEPVIDLKELLTWSAAHVDVLADFDQACEILLWTL
YGIASLYLGSVGNDRARYLARQALHTILLGWRTEPANGRPASS"
2694..3098
/note=".3098"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC68159.1"
/db_xref="GI:29604088"
/translation="MIYHNRYVMKATATPEQIEGVLESWRNQGSRNPAIKGFVWGRN
HGDTYGAVFVEHLDGLFAYLTHPTTYQTDHLGLHLVERLEIFDVSDDDPDLNAK
IQELHRRNELNPQIAGMLADVPTYTSGVDD"
3127..4029
/note="SAV450"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC68160.1"
/db_xref="GI:29604089"
/translation="WAKHVIIGASIGTNNVARLLAEGCESVRIVTRSGSPHEPLVDR
VAADSDRLTELSRGAGVITYCANPPSVTWERLLPPLQTAATAKANADAVLALT
GSLYXGPPQGRMNEHTPMAATGHKRLRRMBOALAGGIRTVVEVRGSDYIGKAN
GIYSLLISALQKGAAMITGLDMPHPTTFNGDMAQALITLASEERAWGRPHWPSA
PAVTIRELARYAAGRPPIKLIQIPRSVTRTAGLIVPIAREMAEMDYQWYAPFMD
ATETTTGLTATDLDTAVRNEVG"
complement(4130..4498)
/gene="bidg"
/complement(4130..4498)
/gene="bidg"
/note="SAV451"
/codon_start=1
/transl_table=11
/product="putative anti-anti-sigma factor"
/protein_id="BAC68161.1"
/db_xref="GI:29604090"
/translation="WTDAPHIDVPHDAAVIALVSGEFDIAAAPAVRARALELIANG
HPDLVDLSGVTFCDSSGALVGIWRCKADGSLTLAAIPDLRLSLLSVTGMDAFL
PAYSDAALAAARQNRRTA"
4635..5156
/note="SAV452"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC68162.1"
/db_xref="GI:29604091"
/translation="WFAVKESVRVVGIEIAATPTGEALBLVLHLGSLRGVRVSSDA
TARPDGLDLAELPARGEWHTTALVRPSVEGEAAGGLPHPGAEPGSGPARRLCWRE
```

```
SAPRIITDIEDVEVHRTLRRSLDGLGAFAHLRPAAPRTHGRRRCPLVHGPPVRPRLAAV
LVHGTAAGPCTPR"
5183..5665
/note="SAV453"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC68163.1"
/db_xref="GI:29604092"
/translation="MVTAVARARASAGWACRARRSSOHSSEGVRRWDEKEFLRTV
AERTGLSRQEAADLTWATLETLARLSPGEARDLITELPEGLAEAVRGTDTDIRERFG
HHDSVVRVAERTMLKEBEADRGVRTVLAVLREAISEKFNLMISQULDFDFOALESAG"
5810..6964
/note="SAV454"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC68164.1"
/db_xref="GI:29604093"
/translation="MIMTADVPTSSGRSVRVALVADPLPSEIARALAPRLPDHLRR
AVGARVRWQVDTMTAPLSVSEQADVSDIAEAVRSHLDGAEDWVAIFLTFDPPRARLYP
ISEVDTALRAALISLPALGVRRLRRVRQAVDVVRELVAHEETPRHPEVIGRPRE
PVPTDRPPQPQRYVVPGLRGHTRLVAGMYANKRMLFGLSLSRALAGVFATVLF
INSATWNLATGALGAQHAVITAGSAVALTLWIIVDHLMWERTGLPARHHPPLYPNLV
TLVTIALVVLAAALPATLVAVSFLLLIDASLRKFIGRQVGVGYLFLAWPITAIAM
IGGAFGTSLRGDEKVSADAAVCRQRQRMLRQMPRPERGERRGPDITDG"
complement(6986..7498)
/note="SAV455"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC68165.1"
/db_xref="GI:29604094"
/translation="WNPYAPASGIEQRPWSWLDYLMWCSDIITADLAEGRLQRPVL
ATVAREAGSERVILAGPAQVTRWALGDGRYMHQNAVAFSPFVLGSMASMATALLG
NSARRAEANDAOQPRCCGARTPGLHVDLVDVHGTMRLMSHMLGTSSTRRPPSSALRALL
ELVDLHVIAL"
7548..8189
/note="SAV456"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC68166.1"
/db_xref="GI:29604095"
/translation="MATLLRRVCPHEGVRIAAVAGFVLSRRPWTASVYRCVAVGSPQS
EDREVEGSTPTVASASSRLRPPCRRLSVVRCPRRAVPSQAVRSRFRGTGTFGLGDLV
VYQRHDLARHVRHVIDDVHPGLRGAAREVRIATHSLGGVIAVDMATANVPLPTES
LVTFGSQATFFHQVPVPGDLLCQMLAFMRKHQEAETGALPPETGFIAHAGQGW"
Alignment Scores:
Pred. No.: 2,91e-19 Length: 299175
Score: 257.00 Matches: 49
Percent Similarity: 71.26% Conservative: 13
Best Local Similarity: 56.32% Mismatches: 25
Query Match: 54.80% Indels: 0
DB: 1 Gaps: 0
US-09-917-376-5 (1-88) x AP005023 (1-299175)
QY 2 SerGlyGlyVallysValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGlnIle 21
Db 105341 TCGGGCGGCTCAAGTCTCTACAGAACCAACACTCTCGGCCACCGACACGCCATC 105400
QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
Db 105401 CGGCCAGGCTTCGATCGTCACACCGGCGAGCGCTCCCTCGACTGTCCAAGCTCAG 105460
QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61
Db 105461 GCGCGCTACTACTTCAACCCGGGACAGCGGCTCGCCCAACCGCTGGTGCAGACTAC 105520
QY 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81
```

DB:	1	2	Gaps:
US-09-917-376-5 (1-88) x AY298814 (1-1957)			
QY	2	SerGlyGlyValLysValGlnTyrIysAsnAspSerAlaProGlyAspAsnGlnIle 21	
DB	1517	ACGGGTGCGCTCGAGGTCTACTACCGCAACACAGCTTTTCGGCGCGCACACAGCCAGATC 1576	
QY	22	LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41	
DB	1577	GGCGCGGGTGGTGGTGGTCAACACCGGAGCAGCAGCTGCAGCTGCAGCTGAGGAA 1636	
QY	42	ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTyr 61	
DB	1637	ATCCACTACTACTTCAACCAACGAGCCCGGC---GGTACCTCCAGCTTCACTTCGCGACTGG 1693	
QY	62	AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81	
DB	1694	GCTCAAGTGGGTGGCGCAACGCTCAACCGCTCTTCACTGCTG-----TCGGCTCCG 1747	
QY	82	ThrAlaAspTyrTyrLeu 87	
DB	1748	GGCGCGGACACCTCCCTG 1765	
RESULT 7			
AF163837		4567 bp	DNA linear BCT 08-FEB-2000
LOCUS			
DEFINITION			Caldibacillus cellulovorans multidomain beta-1,4-mannanase precursor (manA) gene, complete cds; and unknown genes.
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			
misc_binding			
gene			
CDS			
misc_feature			
misc_feature			
ORIGIN			
Alignment Scores:			
Seq. No.:			
Score:			
Percent Similarity:			
Best Local Similarity:			
Query Match:			

DB:	1	2	Gaps:
US-09-917-376-5 (1-88) x AY298814 (1-1957)			
QY	2	SerGlyGlyValLysValGlnTyrIysAsnAspSerAlaProGlyAspAsnGlnIle 21	
DB	1517	ACGGGTGCGCTCGAGGTCTACTACCGCAACACAGCTTTTCGGCGCGCACACAGCCAGATC 1576	
QY	22	LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41	
DB	1577	GGCGCGGGTGGTGGTGGTCAACACCGGAGCAGCAGCTGCAGCTGCAGCTGAGGAA 1636	
QY	42	ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61	
DB	1637	ATCCACTACTACTTCAACCAACGAGCCCGGC---GGTACCCTCCAGCTTCACTTCGCGACTGG 1693	
QY	62	AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81	
DB	1694	GCTCAAGTGGGTGGCGCAACGCTCAACCGCTCTTCACTGCTG-----TCGGCTCCG 1747	
QY	82	ThrAlaAspTyrTyrLeu 87	
DB	1748	GGCGCGGACACCTCCCTG 1765	
RESULT 7			
AF163837		4567 bp	DNA linear BCT 08-FEB-2000
LOCUS			
DEFINITION			Caldibacillus cellulovorans multidomain beta-1,4-mannanase precursor (manA) gene, complete cds; and unknown genes.
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			
misc_binding			
gene			
CDS			
misc_feature			
misc_feature			
ORIGIN			
Alignment Scores:			
Seq. No.:			
Score:			
Percent Similarity:			
Best Local Similarity:			
Query Match:			

CDS	620..3412 /gene="mana" /codon_start=1 /transl_table=11 /product="multidomain beta-1,4-mannanase precursor" /protein_id="AAP22274.1" /db_xref="GI:6651327"	DB: 1 US-09-917-376-5 (1-88) x AF163837 (1-4567)	1	Gaps:	1
sig_peptide	620..718 /gene="mana" /evidence=not_experimental	QY 3 GlyGlyValLysValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGlnIleLys 22 Db 58 GGAGCGCTGCTGTCAGTATCGCGCGGACACGACGCGGGCGCAACACGCTGAAG 117	3237 bp DNA linear BCT 01-DEC-2000 Caldibacillus cellulovorans putative type IIIB cellulose-binding domain gene, partial cds; and beta-1,4-xylanase Xyna precursor (xyna) gene, complete cds.		
mat_peptide	719..3409 /product="multidomain beta-1,4-mannanase" /gene="mana" /note="Region: proline-threonine linker" /evidence=not_experimental	QY 23 ProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrVal 42 Db 118 CCGATTTCGGATTGTAACCGCGGACGTCGAGCGTGCCTCTCGAGCTGACGATT 177	AP200304 3237 bp DNA linear BCT 01-DEC-2000 Caldibacillus cellulovorans putative type IIIB cellulose-binding domain gene, partial cds; and beta-1,4-xylanase Xyna precursor		
misc_feature	1349..1480 /gene="mana" /note="Region: proline-threonine linker" /evidence=not_experimental	QY 43 ArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAla 62 Db 178 CGGTACTGTTACACGGTGGAC---GGGACAAGCGCAGGTGTTCAACTGCGACTGGGG 234	AP200304 AP200304.1 GI:7385018		
misc_feature	1481..1927 /gene="mana" /note="Region: cellulose-binding domain type IIIB" /evidence=not_experimental	QY 63 AlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThr 82 Db 235 CAGTGGGTGTTGCGAACGTGCGCGGAGCTTCGTGAAGCTTTCGACGGCGGACGGGG 294	Caldibacillus cellulovorans Bacteria; Firmicutes; Bacillales; Alicyclobacillaceae; Caldibacillus.		
misc_feature	1928..2032 /gene="mana" /note="Region: proline-threonine linker" /evidence=not_experimental	QY 83 AlaAspThrTyrLeuGln 88 Db 295 GCGGACTACTATATCGAG 312	1 (bases 1 to 3237) Sunna, A., Gibbs, M.D. and Bergquist, P.L. A novel thermostable multidomain 1,4-beta-xylanase from 'Caldibacillus cellulovorans' and effect of its xylan-binding domain on enzyme activity		
misc_feature	2033..2884 /gene="mana" /note="Region: beta-1,4-mannanase catalytic domain" /evidence=experimental		Microbiology 146 (Pt 11), 2947-2955 (2000)		
misc_feature	2885..2959 /gene="mana" /note="Region: proline-threonine linker" /evidence=not_experimental		11065373 2 (bases 1 to 3237) Sunna, A., Gibbs, M.D. and Bergquist, P.L. Direct Submission		
misc_feature	2960..3409 /gene="mana" /note="Region: cellulose-binding domain type IIIB" /evidence=not_experimental		Submitted (29-OCT-1999) Biological Sciences, Macquarie University, North Ryde, Sydney, New South Wales 2109, Australia Location/Qualifiers		
CDS	3413..4400 /note="orf3" /codon_start=1 /transl_table=11 /product="unknown" /protein_id="AAF22275.1" /db_xref="GI:6651328" /translation="MGASGVFSRTAONGYGLKARCKRCVGNLGLHRLFVVDDEF RPFAGARKELGFSRQNLGEAQNFESCTVFQPDKRRIRERNIGOLFVRFOASI GPKLIEAQEQARARKQASDQPLKNVILHPIQHPYGVGERMDPFHKFIEKQKS RSILTRYI"		1..3237 /organism="Caldibacillus cellulovorans" /mol_type="genomic DNA" /db_xref="taxon:74586" <1..279 /note="orf1" /codon_start=1 /evidence=not_experimental /transl_table=11 /product="putative type IIIB cellulose-binding domain" /protein_id="AAP61648.1" /db_xref="GI:7385019" /translation="QVGCNLRGSFVKLTGRTGADYIETFTTSGAGSLAPGASSGD IQVRINKNDWTNTYNEANDYSYDPTKTSFADNRVTLNRNQLIWGVEP"		
Alignment Scores:					
Score:	1.3e-17	Length:	4567		
Percent Similarity:	225.50	Matches:	41		
Best Local Similarity:	67.44%	Conservative:	17		
Query Match:	47.67%	Mismatches:	27		
ORIGIN					
gene	384..3149 /gene="xyna" /gene="xyna" /note="multidomain protein" /codon_start=1 /transl_table=11 /product="beta-1,4-xylanase Xyna precursor"				



terminator		/EC_number="3.2.1.91"		3456..3472		/gene="cel148C"	
ORIGIN		3456..3472		/gene="cel148C"			
Alignment Scores:		9.26e-16		Length: 3509			
Pred. No.:		209.50		Matches: 40			
Percent Similarity:		70.11%		Conservative: 21			
Best Local Similarity:		45.98%		Mismatches: 25			
Query Match:		44.67%		Indels: 1			
DB:		1		Gaps: 1			
US-09-917-376-5 (1-88) x PAE488933 (1-3509)							
Qy	2	SerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle	21				
Db	2879	ACGGGACGCTGGAAGTGCAGTATCGCAATGAGGTCGCTTCAGGTAATGCAATT	2938				
Qy	22	LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr	41				
Db	2939	ACGCCGCAATTTAACTCGAAAATAACAGGTACACGCGCAATTGATCTGAGTAAGGTGAAG	2998				
Qy	42	ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp	61				
Db	2999	GTTTCGTTATTATTTCACGAAGGAC---ACGCAGCGGACATGAGCTTCTGGTGTGACTAT	3055				
Qy	62	AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro	81				
Db	3056	GCACAGCTGGCAGTCCCAATGTGAGCGCAGCTTGTGCTGTAAATCCGGCGAAGGCG	3115				
Qy	82	ThrAlaAspThrTyrLeuGln	88				
Db	3116	ACGGCAGATACGTATCTGGAG	3136				
RESULT 10							
BSP133614							
LOCUS							
DEFINITION							
ACCESSION							
VERSION							
KEYWORDS							
SOURCE							
ORGANISM							
REFERENCE							
AUTHORS							
TITLE							
JOURNAL							
MEDLINE							
PUBMED							
REFERENCE							
AUTHORS							
TITLE							
JOURNAL							
FEATURES							
source							
1..4161							
/organism="Bacillus sp. BP-23"							
/mol_type="genomic DNA"							
/strain="BP-23"							
/db_xref="taxon:89769"							
398..3525							
/gene="celB"							
398..404							
/gene="celB"							
437..442							
/gene="celB"							
514..519							
/gene="celB"							
532..3525							
-35_signal							
-10_signal							
RBS							
CDS							
sig_peptide							
532..633							
/gene="celB"							
ORIGIN							
Alignment Scores:							
Pred. No.:							
Score:							
Percent Similarity:							
Best Local Similarity:							
Query Match:							
DB:							
US-09-917-376-5 (1-88) x BSP133614 (1-4161)							
Qy	2	SerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle	21				
Db	3067	ACAGGTACGCTTCAGGTGCAGTATCGAAGCGGGGTTCCAGCAATTCACGCAATGCAGTT	3126				
Qy	22	LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr	41				
Db	3127	ACTCCGCAATTCATCTGAAGAATACCGCACACAGCGCATTCAGTACGCTGAAA	3186				
Qy	42	ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp	61				
Db	3187	ATTCTGTTATTCATTACCAAGAC---GGCAGGAGGAGCTGTCTCTCTGGTGTGATAT	3243				
Qy	62	AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro	81				
Db	3244	GGCAGGTGGCTCAGCGAAGTACAGGAATGTTTGTAGCAGTGAATCCGCGCAAGGT	3303				
Qy	82	ThrAlaAspThrTyrLeuGln	88				
Db	3304	ACAGCGGATACGTATGTGGAG	3324				
RESULT 11							
E35142							
LOCUS							
DEFINITION							
ACCESSION							
VERSION							
KEYWORDS							
SOURCE							
ORGANISM							
REFERENCE							
1 (bases 1 to 2029)							
E35142							
LOCUS							
DEFINITION							
ACCESSION							
VERSION							
KEYWORDS							
SOURCE							
ORGANISM							
REFERENCE							
1 (bases 1 to 2029)							
E35142							
LOCUS							
DEFINITION							
ACCESSION							
VERSION							
KEYWORDS							
SOURCE							
ORGANISM							
REFERENCE							
1 (bases 1 to 2029)							
E35142							
LOCUS							
DEFINITION							
ACCESSION							
VERSION							
KEYWORDS							
SOURCE							
ORGANISM							
REFERENCE							
1 (bases 1 to 2029)							
E35142							
LOCUS							
DEFINITION							
ACCESSION							
VERSION							
KEYWORDS							
SOURCE							
ORGANISM							
REFERENCE							
1 (bases 1 to 2029)							
E35142							
LOCUS							
DEFINITION							
ACCESSION							
VERSION							
KEYWORDS							
SOURCE							
ORGANISM							
REFERENCE							
1 (bases 1 to 2029)							
E35142							
LOCUS							
DEFINITION							
ACCESSION							
VERSION							
KEYWORDS							
SOURCE							
ORGANISM							
REFERENCE							
1 (bases 1 to 2029)							
E35142							
LOCUS							
DEFINITION							
ACCESSION							
VERSION							
KEYWORDS							
SOURCE							
ORGANISM							
REFERENCE							
1 (bases 1 to 2029)							
E35142							
LOCUS							
DEFINITION							
ACCESSION							
VERSION							
KEYWORDS							
SOURCE							
ORGANISM							
REFERENCE							
1 (bases 1 to 2029)							
E35142							
LOCUS							
DEFINITION							
ACCESSION							
VERSION							
KEYWORDS							
SOURCE							
ORGANISM							
REFERENCE							
1 (bases 1 to 2029)							
E35142							
LOCUS							
DEFINITION							
ACCESSION							
VERSION							
KEYWORDS							
SOURCE							
ORGANISM							
REFERENCE							
1 (bases 1 to 2029)							
E35142							
LOCUS							
DEFINITION							
ACCESSION							
VERSION							
KEYWORDS							
SOURCE							
ORGANISM							
REFERENCE							
1 (bases 1 to 2029)							
E35142							
LOCUS							
DEFINITION							
ACCESSION							
VERSION							
KEYWORDS							
SOURCE							
ORGANISM							
REFERENCE							
1 (bases 1 to 2029)							
E35142							
LOCUS							
DEFINITION							
ACCESSION							
VERSION							
KEYWORDS							
SOURCE							
ORGANISM							
REFERENCE							
1 (bases 1 to 2029)							
E35142							
LOCUS							
DEFINITION							
ACCESSION							
VERSION							
KEYWORDS							
SOURCE							
ORGANISM							
REFERENCE							
1 (bases 1 to 2029)							
E35142							
LOCUS							
DEFINITION							
ACCESSION							
VERSION							
KEYWORDS							
SOURCE							
ORGANISM							
REFERENCE							
1 (bases 1 to 2029)							
E35142							
LOCUS							
DEFINITION							
ACCESSION							
VERSION							
KEYWORDS							
SOURCE							
ORGANISM							
REFERENCE							
1 (bases 1 to 2029)							
E35142							
LOCUS							
DEFINITION							
ACCESSION							
VERSION							
KEYWORDS							
SOURCE							
ORGANISM							
REFERENCE							
1 (bases 1 to 2029)							
E35142							
LOCUS							
DEFINITION							
ACCESSION							
VERSION							
KEYWORDS							
SOURCE							
ORGANISM							
REFERENCE							
1 (bases 1 to 2029)							
E35142							
LOCUS							
DEFINITION							
ACCESSION							
VERSION							
KEYWORDS							
SOURCE							
ORGANISM							
REFERENCE							
1 (bases 1 to 2029)							
E35142							
LOCUS							
DEFINITION							
ACCESSION							
VERSION							
KEYWORDS							
SOURCE							
ORGANISM							
REFERENCE							
1 (bases 1 to 2029)							
E35142							
LOCUS							
DEFINITION							
ACCESSION							
VERSION							
KEYWORDS							
SOURCE							
ORGANISM							
REFERENCE							
1 (bases 1 to 2029)							
E35142							
LOCUS							
DEFINITION							
ACCESSION							
VERSION							
KEYWORDS							
SOURCE							
ORGANISM							
REFERENCE							
1 (bases 1 to 2029)							
E35142							

PN	JP 1999221086-A/45
PD	17-AUG-1999
PF	21-SEP-1998 JP 1998283606
PR	19-SEP-1997 US 08/932571
PI	PAIJI ANDERSON, PETAER L BAGUKUISUTO, ROY M DANIEL, PI
GURAHAMU K FARINTON,	
PI	MORELAND DAVID GIBUSU, HYU MORGAN, DAJAN PURATONIOTSU WILLIAM
PC	C11N15/09, C11N13/386, C12N1/21, C12N9/42// (C12N1/21, C12R1:19), (C12N9/42, C12R1:19), C12N15/00
CC	
PH	Key Location/Qualifiers
FT	source 1. .2029 /organism='Unidentified'.
FEATURES	Location/Qualifiers
source	1. .2029 /organism="unidentified" /mol_type="genomic DNA" /db_xref="taxon:32644"
ORIGIN	
Alignment Scores:	
Pred. No.:	2 6e-14 Length: 2029
Score:	195.50 Matches: 38
Percent Similarity:	60.23% Conservative: 15
Best Local Similarity:	43.18% Mismatches: 34
Query Match:	41.68% Indels: 1
DB:	Gaps: 1
US-09-917-376-5 (1-88) x E35143 (1-2029)	
QY	1 ValSerGlyGlyVallysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db	1 ATGGGAAGTGGTGTGAAGTACTGTACAAAGACAATGAGCAAGTCGAGCAGAGTTCT 60
QY	21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db	61 ATAAGCCCGTGGTTTAAAGATAGTGAATGTGAGGCGAGCAGCAGTGTGTATCTTAGCAGGGTT 120
QY	41 ThrValArgTyrTrpPheThrArgaspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db	121 AGATAAAGATACTGGTGTACACAGTGGATGTGTGCAAGCCACAGAGTCGGTA---TGTGAC 177
QY	61 TtpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db	178 TGGGCACAGATAGGGCGCAGCAATGTGCATTCAATTTTGTGAAGCTTAGCAGCGGAGTG 237
QY	81 ProThrAlaAspThrTyrLeuGln 88
Db	238 AGTGGAGCGGATTAATACCTGGAG 261
RESULT 13	
AF078038S5	6005 bp DNA linear BCT 11-FEB-2001
LOCUS	Caldicellulosiruptor sp. Tok7B.1 glycosyl hydrolase 6 gene, partial
DEFINITION	cds; and CelE gene, complete cds.
ACCESSION	AF078042
VERSION	AF078042.1 GI:12743878
KEYWORDS	5 of 5
SEGMENT	Caldicellulosiruptor sp. Tok7B.1
SOURCE	Caldicellulosiruptor sp. Tok7B.1
ORGANISM	Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae; Caldicellulosiruptor.
REFERENCE	1 (bases 1 to 6005)
AUTHORS	Gibbs,M.D., Reeves,R.A., Farrington,G.K., Anderson,P., Williams,D.P. and Bergquist,P.L.
TITLE	Multidomain and multifunctional glycosyl hydrolases from the extreme thermophile Caldicellulosiruptor isolate Tok7B.1
JOURNAL	Curt. Microbiol. 40 (5), 333-340 (2000)
MEDLINE	20171169
PUBMED	10706665
REFERENCES	2 (bases 1 to 6005)
AUTHORS	Gibbs,M.D., Reeves,R.A., Farrington,G.K., Anderson,P.

Williams, D.P. and Bergquist, P.L.  
Direct Submission  
JOURNAL  
Submitted (27-OCT-1999) Biological Sciences, Macquarie University,  
Sydney, NSW 2109, Australia  
Location/Qualifiers  
FEATURES  
source  
1. .6005  
/organism="Caldicellulosiruptor sp. Tok7B.1"  
/mol\_type="genomic DNA"  
/isolate="Tok7B.1"  
/db\_xref="taxon:80339"  
CDS  
<1. .506  
/note="ORF6; multidomain"  
/codon\_start=3  
/transl\_table=11  
/product="glycosyl hydrolase 6"  
/protein\_id="AAK06393.1"  
/db\_xref="GI:12743884"  
/translation="GQPTWNGTYTGNPNLHVIVDYGTDLGITASLANALYSAAT  
KEYGVSDBAAXNKLKELDRMNLVDRDKGLSAPKRGDYKRFQEVYIPAGWTGKM  
PNGDVKGSGVGFIDIRSKYKQDPDWQKLVSAYNAGEAPEFRYHFWAQCDIAIATY  
BILFGNQ"  
misc\_feature  
1. .503  
/note="Region: family 48 glycosyl hydrolase domain"  
617. .5872  
/note="multidomain glycosyl hydrolase"  
/codon\_start=1  
/evidence="experimental"  
/transl\_table=11  
/product="CelB"  
/protein\_id="AAK06394.1"  
/db\_xref="GI:12743885"  
/translation="MQEMKAIKRVVSVITALLVTLISLCPFGMPVKAYAGTNYNGEA  
LQKIMFYFQMSGLKPSWVRNWRGSDGLDDGVDGLDTGCMHDAGDHFVFNFLPMS  
YSASMLGWAVYDAFKVSKOLEHLNQIEWANDYFVKCHPSKYVYVQVGDPTVDH  
NFWGPAEVMQKRPAYKCDLSPASSVAETAASLAVASVWIKERNKSAKASVLOHAK  
DLFEADTRSDAGTYATGTYTSGFIDDLGAAWVLYIATNDSYTLKAEALMSEY  
ANGNTWTCQDDVRYGTLIMLAKITKRELYKGAVERNLDDHWDRIYTPKGMALYTG  
WGSRYATTAFLACVADWSDCKNTKYLNFPAKSDIQDYALGSGTGRSFVVGFTNY  
PQHPRHNAHSWAMKIPYEHRIHLYGALVGGPDSDDSYNDIDITYQVNEVACDYN  
AGIVGALKALQYLYGEDIPTNDIEIFVESKFGNSQGNPTVETVSYNRT  
GMPRTVDKLSFKYFIDELTELQAGVSDPVKVDYVYLEGKISGPIVWDKKNLYVY  
LVDSGTLKIPGGEVHEHKKQAQKISVPGQYEPDNDPSYKGLTSQLEKNIKIAAYD  
NNNLWGLEPAGATSTPATSTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT  
YKNNETSASTGSIRFWKIVNGSGSSVLSRVKIRYKVTVDGKQPSAVDWAQKIGAS  
NVTFFNLKSLYSGADYLEVFGSGAGLOPKDGTGDIQVRFNKNDNSYNQADDDW  
SWLQSMFNGENAKVTLVYDGLVWGQEPGATPATSTPATPTPTPTPTPTPTPTPTPT  
TPTVTSATPTPATSPVGSYWTPESEYKALKVWANGNLSPTVWLNPKKIENVTG  
TAVDLRSVKRYTIDGATQSVSVASSINPAYIDVRFVKLGANAGGADYIYEIGFK  
SGAGVLAAGQSTKEIRLSIQKSGSNQSDYSVRSANSYIENEKVTGYIDDLVWGR  
EPGRNAQIKWYANGNLGSMTNLNPKIENVTGTAVDLSRVKRYVYTTIDGATQGS  
VSVTSSINPAYIDVRFVKLGANAGADYIYEIGKSGAGVLAAGQSTKEIRLSIQKGS  
GSYNQSDYSVRSATGYIENEKVTGYIDGAIWGREPSRGTKPAGSVTPTPATPTST  
PT  
PARRUGNRLLTYNNWNNNSAGSDWTHSSDDIMCYIMGITGNDKNVPAVSVKFEHQ  
SIKQNASAITLQWYGVYAGKNGTQSVSETPASPRWAEVFKPKGALSQPDVNDNY  
VYMDFFINLYKNGSSATGKIGYILNEDPLMTFTPHRPHQKQVTCSELINKSVE  
LAKVITLDPDAEIPGPAQSGVYLTLDAPQWQKNGHFWLSWYLEOMKASDS  
FGKRLDLVDLHWYPEAQGVVRIQFDGENSTSDVAIRMOAQRPTLDWPTVTKTKG  
QITAGENSWINQWFEYLLPNIKADIDKYPGKLAITEFDYCGKDHISGIALAD  
VLGIFEGVGVYMAAWGSGSYQAQAINIYLNLDGKSGRYGTCVSAETTDVENMPVY  
ASIEGDTSTWAAHILINRYDRKLKAEIKMNNTRVYTGGEIYGFDSSTQIRKMGVLS  
NIQNTTITIEVENLTVYHIVLTSSK"  
725. .2053  
/note="Region: family 9 glycosyl hydrolase domain"  
2054. .2545  
/note="type IIc cellulose binding site"  
2627. .3106  
/note="type IIb cellulose binding site"  
3224. .4195  
/note="type IIb cellulose binding site"  
4298. .5869  
/note="type IIb cellulose binding site"  
2537. .2626  
/note="Region: family 44 glycosyl hydrolase domain"

/note="PT-linker"  
3107. .3223  
/note="PT-linker"  
4196. .4295  
/note="PT-linker"  
ORIGIN  
Alignment Scores:  
Pred. No.: 1.36e-13 Length: 6005  
Score: 194.50 Matches: 38  
Percent Similarity: 61.18% Conservative: 14  
Best Local Similarity: 44.71% Mismatches: 32  
Query Match: 41.47% Indels: 1  
DB: 1 Gaps: 1  
US-09-917-376-5 (1-88) x AF078038S5 (1-6005)  
QY 4 GlyVallyValGlnTyrIlyAsnAsnAspSerAlaProGlyAspAsnGlnIlelySPro 23  
Db 2648 GGTGTGAAGGTACTGTACAAAGAACATGAGACAAGTCGACACACAGGTTCTATAGGCGG 2707  
QY 24 GlyLeuGlnLeuValAsnThrGlySerSerSerValAspLeuSerThrValThrValArg 43  
Db 2708 TGGTTTAAGATAGTGAATGCGAGCAGCAGCAGTGTGATCTTAGCAGGTTAAGATAGA 2767  
QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63  
Db 2768 TACTGGTACACAGTGGATGGTGCACAGCCACAGAGTGGCGTA--TGTCAGCTGGGCACAG 2824  
QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83  
Db 2825 ATAGGGCGAAGCAATGTGACATTTTGTGAAGCTTAGCAGCGAGTGGAGCGG 2884  
QY 84 AspThrTyrLeuGln 88  
Db 2885 GATTATACCTGGAG 2899  
RESULT 14  
E35100  
LOCUS Truncated cellulase composition.  
DEFINITION E35100  
ACCESSION E35100  
VERSION E35100.1 GI:13018925  
KEYWORDS JP 1999221086-A/2.  
SOURCE unidentified  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 6416)  
AUTHORS Paiji,A., Petaer,L.B., Roy,M.D., Gurahamu,K.F., Moreland,D.G.,  
Hyu,M. and Daian,P.W.  
TITLE Truncated cellulase composition  
JOURNAL Patent: JP 1999221086-A 2 17-AUG-1999;  
COMMENT CLARIANT INTERNATIONAL LTD  
OS Unknown  
PN JP 1999221086-A/2  
PD 17-AUG-1999  
PR 21-SEP-1998 JP 1998283606  
PR 19-SEP-1997 US 08/932571  
PI PAJJI ANDERSON, PETAER L BAGUKISUTO, ROY M DANIEL, PI  
GURAHAMU K FARINTON,  
PI MORELAND DAVID GIBUSU, HYU MORGAN, DAIAN PURATONOTISU WILLIAM  
PC C12N15/09, C12N13/386, C12N1/21, C12N9/42// (C12N1/21, C12R1:19), PC  
(C12N9/42, C12R1:19), C12N15/00  
CC Key Location/Qualifiers  
FH source 1. .6416  
FT /organism="Unknown".  
FEATURES  
source  
1. .6416  
/organism="unidentified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"  
ORIGIN





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 12, 2004, 03:07:10 ; Search time 122.068 Seconds  
(without alignments)  
3062.577 Million cell updates/sec

Title: US-09-917-376-5

Perfect score: 469

Sequence: 1 VSGGVKQVKNNDSPGDNQ.....IRASFGSVNPATPTADTYLQ 88

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09917376/runat\_11052004\_114534\_28231/app\_query.fasta\_1.1429  
-DB=N Geneseq\_29Jan04 -QPMF=fastcap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0  
-LOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09917376 @CGN 1 1 758 @runat\_11052004\_114534\_28231 -NCPU=6 -ICPU=3  
-NO\_WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq\_29Jan04.\*

1: Geneseqn1980s.\*  
2: Geneseqn1990s.\*  
3: Geneseqn2000s.\*  
4: Geneseqn2001as.\*  
5: Geneseqn2001bs.\*  
6: Geneseqn2002s.\*  
7: Geneseqn2003as.\*  
8: Geneseqn2003bs.\*  
9: Geneseqn2003cs.\*  
10: Geneseqn2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	469	100.0	2289	7 ABZ77633	Abz77633 Nucleotid
2	466	99.4	3687	7 ABZ77634	Abz77634 Nucleotid
3	462	98.5	2869	7 ABZ77632	Abz77632 Nucleotid
4	462	98.5	2869	9 ADD22922	Add22922 Acidothe
5	436	93.0	3365	7 ABZ76162	Abz76162 A. cellu
6	231.5	49.4	2600	2 AAQ15178	Aaq15178 Portion o
7	195.5	41.7	2029	2 AAQ15178	Aax55660 DNA seque
8	195.5	41.7	2029	6 AAD26568	Aad26568 Active ce

9	194.5	41.5	6415	2 AAX55662	Aax55662 DNA seque
10	194.5	41.5	6416	6 AAD26566	Aad26566 Active ce
11	192.5	41.0	11706	2 AAX55661	Aax55661 DNA seque
12	192.5	41.0	11707	6 AAD26525	Aad26525 Active ce
13	188	40.1	1434	6 AAL41028	Aal41028 CMCase ge
14	188	40.1	1488	6 AAL41025	Aal41025 CMCase ge
15	188	40.1	2510	6 ABK53202	Abk53202 Bacillus
16	186	39.7	2175	2 AAQ49820	Aaq49820 NK-1 cell
17	182.5	38.9	2977	3 AAQ13001	Aaq13001 Endo1 gen
18	165	35.2	1438	3 AAZ45336	Aaz45336 DNA encod
19	165	35.2	1482	2 AAX90978	Aax90978 DNA encod
20	165	35.2	1482	2 AAZ31562	Aaz31562 Pectate 1
21	165	35.2	5562	2 AAT86625	Aat86625 C. thermo
22	157	33.5	1314	6 ABK73393	Abk73393 Bacillus
23	155.5	33.2	4059	5 AAH75059	Aah75059 Nucleotid
24	150.5	32.1	4056	5 AAH75060	Aah75060 Nucleotid
25	130.5	27.8	4452	7 ACA27429	Ac27429 Prokaryot
26	126.5	27.0	1303	9 ADC27473	Adc27473 Fusion pr
27	126.5	27.0	1747	9 ADC27474	Adc27474 Fusion pr
28	126.5	27.0	3489	9 ADC27475	Adc27475 Fusion pr
29	120.5	25.7	1775	2 AAQ13003	Aaq13003 Endo3 gen
30	114	24.3	486	2 AAQ72917	Aaq72917 Cellulose
31	114	24.3	486	2 AAV74072	Aav74072 C. cellu
32	114	24.3	499	2 AAX24930	Aax24930 Clostridi
33	114	24.3	507	5 AAD11042	Aad11042 Clostridi
34	114	24.3	768	4 AAF86248	Aaf86248 DNA seque
35	114	24.3	768	6 ABK52403	Abk52403 E. coli c
36	114	24.3	984	5 AAD11046	Aad11046 Chimeric
37	114	24.3	1030	5 AAD11044	Aad11044 Clostridi
38	114	24.3	1288	5 AAD11045	Aad11045 Clostridi
39	113	24.1	573	5 AAD11043	Aad11043 Clostridi
40	113	24.1	1146	2 AAZ27550	Aaz27550 Gausia 1
41	113	24.1	1146	6 AAD22201	Aad22201 Gausia s
42	78.5	16.7	764	3 AAF12881	Aaf12881 Aspergill
43	78.5	16.7	3147	7 ACA25846	Aca25846 Prokaryot
44	77	16.4	8107	3 AAZ92616	Aaz92616 Genomic D
45	76	16.2	96597	8 ADA02720	Ada02720 Human NFA

#### ALIGNMENTS

RESULT 1

ABZ77633

ID ABZ77633 standard; DNA; 2289 BP.

XX ABZ77633;

XX DT 03-JUN-2003 (first entry)

XX DE Nucleotide sequence of the ManA polypeptide.

XX KW ManA; mannanase A; glycoside hydrolase; enzyme; hemicellulose; sugar;  
XX food; feed; paper pulp; biofuel; mannanase; gene; ss.

XX OS Acidothermus cellulolyticus.

XX PH Key Location/Qualifiers

XX CDS 1..2289

XX FT /\*tag= a

XX FT /product= "ManA"

XX PN WO2003012110-A1.

XX PD 13-FEB-2003.

XX PF 28-JUL-2001; 2001WO-US023819.

XX PR 28-JUL-2001; 2001WO-US023819.

XX PA (WIDE ) MIDWEST RES INST.

XX PI Ding S, Adney WS, Vinzant TB, Himmel ME;



DE Nucleotide sequence of the avicelase AvIII.

KW Avicelase; AvIII; glycoside hydrolase; enzyme; cellulase; biofuel;  
 KW detergent; pulp processing; paper processing; feed processing; textile;  
 KW cellulose; gene; ss.

XX  
 OS Acidothermus cellulolyticus.

XX  
 FH Key Location/Qualifiers  
 FT CDS 1..2869  
 FT /tag= a  
 FT /partial  
 FT /product= "AvIII"  
 FT /transl\_except= (pos:2869,aa:Xaa)  
 FT /note= "Xaa is an unspecified residue"

XX  
 PN WO2003012090-A2.

XX  
 PD 13-FEB-2003.

XX  
 PF 28-JUL-2001; 2001WO-US023818.

XX  
 PR 28-JUL-2001; 2001WO-US023818.

XX  
 PA (MIDE ) MIDWEST RES INST.

XX  
 PI Ding S, Adney WS, Vinzant TB, Himmel ME;  
 XX  
 DR WPI; 2003-248177/24.  
 DR P-PSDB; ABP73015.

XX  
 PT New thermostable AvIII peptide from Acidothermus cellulolyticus, useful  
 PT for degradation of cellulose or in generating anti-AvIII antibodies for  
 PT purifying recombinant AvIII polypeptides from genetically engineered  
 PT host cells.

XX  
 PS Claim 3; Page 24; 44pp; English.

XX  
 CC The present sequence encodes a thermostable avicelase polypeptide,  
 CC designated AvIII. AvIII is a member of the glycoside hydrolase family  
 CC of enzymes, and is a cellulase. AvIII is useful in the conversion of  
 CC biomass to biofuels and biofuel additives. It may be useful in the  
 CC production of detergents, pulp and paper processing, food and feed  
 CC processing and in textile processes. The thermostable AvIII peptide is  
 CC useful in the degradation of cellulose, and in generating specific anti-  
 CC AvIII antibodies that are useful in purifying recombinant AvIII  
 CC polypeptides from genetically engineered host cells, in detecting AvIII  
 CC polypeptide expression, as well as a reagent tool for characterizing the  
 CC molecular actions of the polypeptide. The AvIII polynucleotide is useful  
 CC as a source of probes or primers in various diagnostic assays

XX  
 SQ Sequence 2869 BP; 546 A; 857 C; 900 G; 565 T; 0 U; 1 Other;

Alignment Scores:  
 Pred. No.: 4.09e-47 Length: 2869  
 Score: 462.00 Matches: 86  
 Percent Similarity: 100.00% Conservative: 2  
 Best Local Similarity: 97.73% Mismatches: 0  
 Query Match: 98.51% Indels: 0  
 DB: 7 Gaps: 0

US-09-917-376-5 (1-88) x AB277632 (1-2869)

QY 1 ValSerGlyValLeuValGlnTyrLysAsnAspSerAlaProGlyAspGln 20  
 DB 2605 GTGTCGGTGGGTGAGGTGCAGTATAGAAATATGATTCGGCGCGGTGATAATCAG 2664

QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40  
 DB 2665 ATCAAGCCGGGTTTGACAGTGGTGAATACCGGTCGTGCGGTGATTTGTCGACGGTG 2724

QY 41 ThrValArgTyrTyrPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
 DB

Db 2725 ACGTGGGTACTGGTTTACCCGGGATGGTGGCTCGTGCACACTGGTGTAACCTGTGAC 2784

QY 61 TTPAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80  
 DB 2785 TGGGGCGCGATCGGGTGTGGGAATATCCGCCCTCGTTCGGTTCGGTGAACCCCGCGACG 2844

QY 81 ProThrAlaAspThrTyrLeuGln 88  
 DB 2845 CCGACGGCGGACACCTACCTGCAG 2868

RESULT 4  
 ADD22922  
 ID ADD22922 standard; DNA; 2869 BP.  
 XX  
 AC ADD22922;  
 XX  
 DT 15-JAN-2004 (first entry)  
 XX  
 DE Acidothermus cellulolyticus avicelase AvIII DNA.  
 KW AvIII; cellulose reduction; agricultural biomass; municipal solid waste;  
 KW glycoside hydrolase; avicelase; ds; gene.  
 XX  
 OS Acidothermus cellulolyticus.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..2868  
 FT /tag= a  
 FT /product= "AvIII"  
 XX  
 PN US2003108988-A1.  
 XX  
 PD 12-JUN-2003.  
 XX  
 PF 18-OCT-2002; 2002US-00155400.  
 XX  
 PR 28-JUL-2001; 2001US-00917376.  
 XX  
 PA (DING/) DING S.  
 PA (ADNE/) ADNEY W S.  
 PA (VINZ/) VINZANT T B.  
 PA (HIMM/) HIMMEL M E.  
 XX  
 PI Ding S, Adney WS, Vinzant TB, Himmel ME;  
 XX  
 DR WPI; 2003-810853/76.  
 DR P-PSDB; ADD22921.  
 XX  
 PT New isolated thermal tolerant avicelase polynucleotide useful for  
 PT detection of a polynucleotide encoding AvIII and for reducing cellulose  
 PT in a starting material, e.g. municipal solid waste.  
 XX  
 PS Claim 17; SEQ ID NO 2; 29pp; English.

XX  
 CC The invention relates to an isolated polynucleotide molecule encoding a  
 CC thermostable AvIII polypeptide. The polynucleotide is useful for  
 CC detection of a polynucleotide encoding AvIII. The polynucleotide is  
 CC useful for reducing cellulose in a starting material which involves  
 CC administering to the starting material, e.g. agricultural biomass or  
 CC municipal solid waste, a polypeptide molecule of the polynucleotide. The  
 CC method further comprises administering a second polypeptide molecule  
 CC chosen from the glycoside hydrolase family of proteins. The present  
 CC sequence represents DNA encoding the Acidothermus cellulolyticus  
 CC avicelase AvIII.

XX  
 SQ Sequence 2869 BP; 546 A; 857 C; 900 G; 565 T; 0 U; 1 Other;

Alignment Scores:  
 Pred. No.: 4.09e-47 Length: 2869  
 Score: 462.00 Matches: 86  
 Percent Similarity: 100.00% Conservative: 2  
 Best Local Similarity: 97.73% Mismatches: 0  
 Query Match: 98.51% Indels: 0

DB: 9 Gaps: 0

US-09-917-376-5 (1-88) x ADD22922 (1-2869)

QY 1 ValSerGlyGlyValGlnTyrLysAsnAspSerAlaProGlyAspGln 20

Db 2605 GTGTCGGTGGGTGAAGTGCAGTATAGATAATGATTCGGCGCGGTGATATCAG 2664

QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40

Db 2665 ATCAAGCCGGTTTCAGGTGCTGAATACCGGGTCTGCTCGGTGGATTGTTCGACGGTG 2724

QY 41 ThrValArgTyrTrpPheThrArgSerGlyGlySerSerThrLeuValTyrAsnCysAsp 60

Db 2725 ACGGTGCAGTACTGGTTTACCGCGGATGCTGGCTCGTCACACTGCTGTGAC 2784

QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80

Db 2785 TGGGGCGGATCGGTGTGGGAATATCCGGCTCGTTCCGGTCGGTGAACCGCGGACG 2844

QY 81 ProThrAlaAspThrTyrLeuGln 88

Db 2845 CCGACGGCGGACCTACTGTGAC 2868

RESULT 5

ID ABZ76162 standard; DNA; 3365 BP.

XX AC ABZ76162;

XX DT 29-MAY-2003 (first entry)

XX DE A. cellulolyticus Gux1 protein encoding DNA.

KW Gux1; thermal tolerant; exoglucanase; glycoside hydrolase; cellulose;

KW biofuel; detergent; pulp; paper processing; feed processing; textile;

KW cellulase; gene; ds.

XX OS Acidothermus cellulolyticus.

XX FH Key Location/Qualifiers

FT CDS 1..3365

FT /tag= a

FT /product= "Gux1"

FT /transl\_except= (pos: 682..683, aa: Pro)

FT /note= "this codon has an apparent one nucleotide

FT basepair deletion which alters the reading frame"

XX PN WO2003012095-A1.

XX PD 13-FEB-2003.

XX PF 28-JUL-2001; 2001WO-US023820.

XX PR 28-JUL-2001; 2001WO-US023820.

XX PA (WIDE ) MIDWEST RES INST.

XX PI Adney WS, Ding S, Vinzant TB, Himmel ME, Decker SR;

PI Lantz Mccarter S;

XX WPI; 2003-300494/29.

DR P-PSDB; ABP71656.

XX New thermal tolerant Gux1 peptide having specified amino acid sequence,

PT useful in the degradation of cellulose to biofuels.

XX Disclosure; Page 22-23; 44pp; English.

XX The invention relates to a thermal tolerant Gux1 peptide from A.

CC cellulolyticus. The Gux1 exoglucanase is a member of the glycoside

CC hydrolase family and comprises a catalytic domain GH48, carbohydrate

CC binding domain type III, and a carbohydrate binding domain type II. The

CC polypeptide is useful in the degradation of cellulose into biofuel, or

CC for conversion of biomass to biofuel additives. It is used in detergents,

CC pulp and paper processing, food and feed processing, and in textile

CC processing. It can also be used alone or in combination with other

CC cellulase or glycoside hydrolases. The novel polypeptide generates

CC alternative cellulase enzymes capable of assisting in the commercial-

CC scale processing of cellulose to sugar for use in biofuel production. The

CC present sequence represents a A. cellulolyticus Gux1 cellulase encoding

XX DNA

SQ Sequence 3365 BP; 669 A; 1117 C; 1041 G; 538 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 8.56e-44 Length: 3365

Score: 436.00 Matches: 81

Percent Similarity: 97.62% Conservative: 1

Best Local Similarity: 96.43% Mismatches: 2

Query Match: 92.96% Indels: 0

DB: 7 Gaps: 0

US-09-917-376-5 (1-88) x ABZ76162 (1-3365)

QY 5 ValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysProGly 24

Db 112 CTCAAAGCCGAGTATAAGAAACAATGATTCGGCGCGGAGTACAAACCGGT 171

QY 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArgTyr 44

Db 172 CTCAGTTGGTGAATACCGGGTCTGCTCGGTGGATTGTTCGCGGTGACGGTCCGGTAC 231

QY 45 TrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAlaMet 64

Db 232 TGGTTCACCGGGATGCTGGTCTGCACACTGCTGTACAACTGTGCTGGCGCGGATG 291

QY 65 GlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAlaAsp 84

Db 292 GGGTGTGGGAATATCCGGCTCGTTGGTTCGGTGAACCGCGGACCGCGGAC 351

QY 85 ThrTyrLeuGln 88

Db 352 ACCTACTGCGAG 363

RESULT 6

AAQ15178

ID AAQ15178 standard; DNA; 2600 BP.

XX AC AAQ15178;

XX DT 27-AUG-2003 (revised)

DT 25-MAR-2003 (revised)

DT 17-MAR-1992 (first entry)

XX DE Portion of pAEC-1 contg. cellulase AE-1 gene.

XX KW Detergents; pharmaceuticals; deinking; carboxymethylcellulose; ss.

XX OS Aeromonas sp.

XX FH Key Location/Qualifiers

FT CDS 192..2540

FT /tag= a

FT /product= "cellulase AE-1"

XX PN JP03251174-A.

XX PD 08-NOV-1991.

XX PF 28-FEB-1990; 90JP-00045465.

XX PR 28-FEB-1990; 90JP-00045465.

XX PA (OJIP ) OJI PAPER CO.

DR WPI: 1991-373412/51.  
 DR P-PSDB; AAR15625.  
 XX  
 PT Cellulase AE-1 for e.g. mfr. of pharmaceuticals and foodstuffs - of opt.  
 PT pH when carboxymethylcellulose is used as substrate.  
 XX  
 XX Claim 3; Fig 2; 8pp; Japanese.  
 PS  
 CC The plasmid, pAEC 1, was prepd. ligating chromosomal DNA contg. the gene  
 CC (obtd. from Aeromonas) into pUC18. The gene can be used to produce  
 CC recombinant enzyme which is used for the effective utilis- ation of  
 CC biomass resources and the mfr. of pharmaceuticals and foodstuffs, and  
 CC also for the detergent and denking of waste paper. (Updated on 25-MAR-  
 CC 2003 to correct PA field.) (Updated on 27-AUG-2003 to correct OS field.)  
 XX  
 SQ Sequence 2600 BP; 687 A; 649 C; 700 G; 564 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 1.62e-18 Length: 2600  
 Score: 231.50 Matches: 42  
 Percent Similarity: 66.67% Conservative: 16  
 Best Local Similarity: 48.28% Mismatches: 28  
 Query Match: 49.36% Indels: 1  
 DB: 2 Gaps: 1

US-09-917-376-5 (1-88) x AAQ15178 (1-2600)

QY 2 SerGlyGlyValLeuValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 21  
 Db 2082 AGCGGGGATTTGGCGTTCAGTACAAAACCGGAGATGCAATGCGCGCAACAGTTC 2141

QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41  
 Db 2142 AAGCTCTATTCATATCGTGAACAGGGGCTGCGCAGTGCCTTTAAGCGAGCTGAGT 2201

QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61  
 Db 2202 CTGCGATATTATTTCCACAGCTGAC--GGCAATGACCAACTGCAATACAAATTTGACTGG 2258

QY 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81  
 Db 2259 GCCATGTAGGTGCTCCAACTGAACGGGGCTTTCTGTAATAATGAATCCGGCAAAGCG 2318

QY 82 ThrAlaAspThrTyrLeuGln 88  
 Db 2319 AACGCCGATACCTACTTTGGAG 2339

RESULT 7  
 ID AAX55660  
 XX AAX55660 standard; DNA; 2029 BP.  
 AC AAX55660;  
 XX  
 DT 30-JUL-1999 (first entry)  
 DE  
 XX DNA sequence encoding truncated cellulase Cel E3/B5.  
 KW Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;  
 KW Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;  
 KW cotton-containing fabric; stonewashing; ss.  
 XX  
 OS Unidentified.  
 XX  
 PN EP921188-A2.  
 XX  
 PD 09-JUN-1999.  
 XX  
 PF 15-SEP-1998; 98EP-00810919.  
 XX  
 PR 19-SEP-1997; 97US-00932571.  
 XX  
 PA (CLRN ) CLARIANT FINANCE BVI LTD.  
 XX

PI Anderson P, Bergquist PL, Daniels RM, Farrington GK, Gibbs MD;  
 PI Morgan H, Williams DP;  
 XX  
 DR WPI: 1999-315403/27.  
 DR P-PSDB; AAY13494.  
 XX  
 PT New truncated cellulase proteins, useful in detergents and for producing  
 PT 'stonewashed' denim.  
 PS  
 XX Disclosure; Page 41-42; 65pp; English.  
 XX  
 CC The invention relates to a recombinant cellulase active protein free of  
 CC proteinases of native thermophilic and alkalophilic origin, comprising  
 CC the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3,  
 CC Cel 6 or Cel E3/B5, or a stability region from one of the defined full-  
 CC length sequences, or functional equivalents. Cel B5 extends from amino  
 CC acid A1011 to P1424 or K1425 or N1426, and Cel B4/5 extends from amino  
 CC acid K635 to N1426 in the sequence shown in AAY13492; Cel E1 extends from  
 CC amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel E1/2/3  
 CC extends from Y39 to G812, Cel E6 extends from amino acid V1233 to K1751  
 CC and the stability region extends from amino acid E482 to G635 in the  
 CC sequence shown in AAY13493; Cel E3/B5 is shown in AAY13494. The new  
 CC enzymes are useful in laundry detergent compositions to prevent or remove  
 CC staining, backstaining or graying, for use on cellulosic materials  
 CC including cotton-containing fabrics. They are especially useful for  
 CC preventing redeposition of colorant during stonewashing, and for  
 CC processing of textiles where cellulose breakdown is required. The new  
 CC truncated enzymes show reduced redeposition of dye compared to using non-  
 CC truncated cellulase compositions  
 XX  
 SQ Sequence 2029 BP; 650 A; 329 C; 520 G; 530 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 3.53e-14 Length: 2029  
 Score: 195.50 Matches: 38  
 Percent Similarity: 60.23% Conservative: 15  
 Best Local Similarity: 43.18% Mismatches: 34  
 Query Match: 41.68% Indels: 1  
 DB: 2 Gaps: 1

US-09-917-376-5 (1-88) x AAX55660 (1-2029)

QY 1 ValSerGlyGlyValLeuValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20  
 Db 1 ATGGGAAGTGGTGTGAAGGTACTGTACAGAACATGACAGCAAGTGGCAGCAGGTCT 60

QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40  
 Db 61 ATAAGGCCGCTGGTTTAAGATAGTGAATGGAGCGCAGCAGTGTTCATCTTAGCAGGGTT 120

QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
 Db 121 AAGATAAGATACCTGTTACACAGCTGGATGGTGTGACAGCCACAGAGTGGCGTA---TGTGAC 177

QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80  
 Db 178 TGGGCACAGATAGGGCAAGCAATGTGACATTCATTTTGTGAAGCTTAGCAGCGGAGTG 237

QY 81 ProThrAlaAspThrTyrLeuGln 88  
 Db 238 AGTGGAGCGGATTATTACCTGGAG 261

RESULT 8  
 ID AAD26568  
 XX AAD26568 standard; DNA; 2029 BP.  
 AC AAD26568;  
 XX  
 DT 26-MAR-2002 (first entry)  
 DE  
 XX Active cellulase hybrid protein, E3/B5 DNA.  
 XX  
 KW Active cellulase protein; alkalophilic; textile processing; proteinase;

KW detergent additive; stonewashed appearance; cotton-containing denim;  
 KW CelB5; thermophilic; commercial detergent; E3/B5 hybrid protein; ds.  
 XX Unidentified.  
 OS  
 XX  
 XX  
 FH Key Location/Qualifiers  
 FT 1..1851  
 FT CDS  
 FT /tag= a  
 FT /product= "E3/B5 hybrid protein"  
 XX  
 XX US62943366-B1.  
 XX  
 XX PD 25-SEP-2001.  
 XX  
 XX PF 19-AUG-1998; 98US-00136574.  
 XX  
 XX PR 19-SEP-1997; 97US-00932571.  
 XX  
 XX PA (CLRN ) CLARIANT FINANCE BVI LTD.  
 XX  
 XX PI Farrington GK, Anderson P, Bergquist P, Daniels R, Gibbs MD;  
 XX Morgan H, Williams DP;  
 XX  
 XX DR WPI; 2002-081780/11.  
 XX P-PSDB; AAE16325.  
 XX  
 XX New cellulase active protein, useful in textile processing or commercial  
 PT detergents, e.g. for improving the feel or appearance of cotton-  
 PT containing fabrics, is stable under conditions of alkaline pH and  
 PT elevated temperatures.  
 XX  
 XX PS Disclosure; Col 71-74; 61pp; English.  
 XX  
 XX The present invention relates to a cellulase active protein, which is  
 CC substantially free of proteinases of native thermophilic and  
 CC alkalophilic origin, where the cellulase active protein consists of the  
 CC CelB5 amino acid sequence. The cellulase active protein is useful for  
 CC treating cellulosic materials including cotton-containing fabrics, as  
 CC detergent additives. The cellulase active protein is also useful for  
 CC improving the feel and/or appearance of cotton-containing fabrics, for  
 CC removing surface fibers from cotton-containing knits or for imparting  
 CC stonewashed appearance to cotton-containing denims. The present proteins  
 CC are stable under condition of alkaline pH and elevated temperatures, thus  
 CC suitable for textile processing and in commercial detergents. The present  
 CC sequence is E3/B5 hybrid protein DNA  
 XX  
 XX SQ Sequence 2029 BP; 650 A; 329 C; 520 G; 530 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 3.53e-14 Length: 2029  
 Score: 195.50 Matches: 38  
 Percent Similarity: 60.23% Conservatives: 15  
 Best Local Similarity: 43.18% Mismatches: 34  
 Query Match: 41.68% Indels: 1  
 DB: 6 Gaps: 1  
 US-09-917-376-5 (1-88) x AAD26568 (1-2029)  
 QY 1 ValSerGlyValValysValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGln 20  
 Db 1 ATGGGAAGTGGTGTGAAGTACTGTACAAAGCAATGAGCAACGTCGAGCACAGGTTCT 60  
 QY 21 IleLysProGlyLeuGlnLeuValThrGlySerSerValAspLeuSerThrVal 40  
 Db 61 ATAAGCCCGTGTAAAGATAGTGAATGAGGAGCAGCAGTGTGATCTTAGCAGGGTT 120  
 QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
 Db 121 AGATAAGATATCTGTACACAGTGGTGTGACAGCCACAGATGCCGTA---TGTGAC 177  
 QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80  
 Db 178 TGGGCACAGATAGGGCAAGCAATGTGACATTCATTTTGTGAAGCTTAGCAGCGAGTG 237

QY 81 ProThrAlaAspThrTyrLeuGln 88  
 Db 238 AGTGGAGCGGATTATTACTGGAG 261  
 RESULT 9  
 AAX55662  
 ID AAX55662 standard; DNA; 6415 BP.  
 XX  
 XX AAX55662;  
 XX  
 DT 30-JUL-1999 (first entry)  
 XX  
 XX DNA sequence encoding truncated cellulases.  
 XX  
 XX Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;  
 KW Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;  
 KW cotton-containing fabric; stonewashing; ss.  
 XX  
 XX Unidentified.  
 XX  
 XX PN BP921188-A2.  
 XX  
 XX PD 09-JUN-1999.  
 XX  
 XX PF 15-SEP-1998; 98EP-00810919.  
 XX  
 XX PR 19-SEP-1997; 97US-00932571.  
 XX  
 XX PA (CLRN ) CLARIANT FINANCE BVI LTD.  
 XX  
 XX PI Anderson P, Bergquist PL, Daniels RM, Farrington GK, Gibbs MD;  
 XX Morgan H, Williams DP;  
 XX  
 XX DR WPI; 1999-315403/27.  
 XX P-PSDB; AAY13493.  
 XX  
 XX New truncated cellulase proteins, useful in detergents and for producing  
 PT 'stonewashed' denim.  
 XX  
 XX PS Disclosure; Page 24-25; 65pp; English.  
 XX  
 XX The invention relates to a recombinant cellulase active protein free of  
 CC proteinases of native thermophilic and alkalophilic origin, comprising  
 CC the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3  
 CC Cel 6 or Cel E3/B5, or a stability region from one of the defined full-  
 CC length sequences, or functional equivalents. Cel B5 extends from amino  
 CC acid A1011 to P1424 or K1425 or N1426, and Cel B4/5 extends from amino  
 CC acid K635 to N1426 in the sequence shown in AAY13492; Cel E1 extends from  
 CC amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel E1/2/3  
 CC and the stability region extends from amino acid E482 to G635 in the  
 CC sequence shown in AAY13493; Cel E3/B5 is shown in AAY13494. The new  
 CC enzymes are useful in laundry detergent compositions to prevent or remove  
 CC staining, backstaining or graying, for use on cellulosic materials  
 CC including cotton-containing fabrics. They are especially useful for  
 CC preventing redeposition of colorant during stonewashing, and for  
 CC processing of textiles where cellulose breakdown is required. The new  
 CC truncated enzymes show reduced redeposition of dye compared to using non-  
 CC truncated cellulase compositions  
 XX  
 XX SQ Sequence 6415 BP; 2067 A; 1082 C; 1689 G; 1541 T; 0 U; 36 Other;  
 Alignment Scores:  
 Pred. No.: 2.05e-13 Length: 6415  
 Score: 194.50 Matches: 38  
 Percent Similarity: 61.18% Conservatives: 14  
 Best Local Similarity: 44.71% Mismatches: 32  
 Query Match: 41.47% Indels: 1  
 DB: 2 Gaps: 1  
 US-09-917-376-5 (1-88) x AAX55662 (1-6415)

QY 4 GlyVallysValGlnTyrIysAsnAspSerAlaProGlyAspAsnGlnIleIysPro 23  
 Db 2664 GGTGTGAAGTACTGTACAAACAATGAGCAAGTGCAGCACAGGTTCTATAAGCGC 2723  
 QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43  
 Db 2724 TGGTTTAAGTAGTAAGTGGAGGCAGCAGTGTGATCTTAGCAGGTTAAGATAAGA 2783  
 QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaLa 63  
 Db 2784 TACTGTACACAGTGGATGGTGACAGCCACAGAGTGGCGTA--TGTGACTGGGCACAG 2840  
 QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83  
 Db 2841 ATAGGGCAAGCAATGTGACATTCAATTTGTGAAGCTTAGCAGCGGAGTGAGTGGAGCG 2900  
 QY 84 AspThrTyrLeuGln 88  
 Db 2901 GATTATTACCTGGAG 2915

RESULT 10  
 AAD26526  
 ID AAD26526 standard; DNA; 6416 BP.  
 XX  
 AC AAD26526;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Active cellulase protein, cele gene.  
 XX  
 KW Active cellulase protein; alkaliphilic; textile processing; proteinase;  
 KW detergent additive; stonewashed appearance; cotton-containing denim;  
 KW CelB5; thermophilic; commercial detergent; cele gene; ds.  
 XX  
 OS Unidentified.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 634..5889  
 FT /tag= a  
 FT /product= "cele protein"  
 FT misc\_feature 748..2538  
 FT /tag= c  
 FT /product= "DNA encoding E1/2 protein"  
 FT misc\_feature 748..2076  
 FT /tag= b  
 FT /product= "DNA encoding E1 protein"  
 XX  
 PN US6294366-B1.  
 XX  
 PD 25-SEP-2001.  
 XX  
 PF 19-AUG-1998; 98US-00136574.  
 XX  
 PR 19-SEP-1997; 97US-00932571.  
 XX  
 PA (CLRN ) CLARIANT FINANCE BVI LTD.  
 XX  
 PI Farrington GK, Anderson P, Bergquist P, Daniels R, Gibbs MD;  
 PI Morgan H, Williams DP;  
 DR WPI: 2002-081780/11.  
 DR P-PSDB; AA616324.  
 XX  
 PT New cellulase active protein, useful in textile processing or commercial  
 PT detergents, e.g. for improving the feel or appearance of cotton-  
 PT containing fabrics, is stable under conditions of alkaline pH and  
 PT elevated temperatures.  
 XX  
 PS Disclosure; Col 37-44; 61pp; English.  
 XX  
 CC The present invention relates to a cellulase active protein, which is  
 CC substantially free of proteinases of native thermophilic and  
 CC alkaliphilic origin, where the cellulase active protein consists of the

CC CelB5 amino acid sequence. The cellulase active protein is useful for  
 CC treating cellulosic materials including cotton-containing fabrics, as  
 CC detergent additives. The cellulase active protein is also useful for  
 CC improving the feel and/or appearance of cotton-containing fabrics, for  
 CC removing surface fibers from cotton-containing knits or for imparting  
 CC stonewashed appearance to cotton-containing denims. The present proteins  
 CC are stable under condition of alkaline pH and elevated temperatures, thus  
 CC suitable for textile processing and in commercial detergents. The present  
 CC sequence is cele gene  
 XX  
 SQ Sequence 6416 BP; 2068 A; 1082 C; 1689 G; 1541 T; 0 U; 36 Other;

Alignment Scores:  
 Pred. No.: 2, 05e-13 Length: 6416  
 Score: 194.50 Matches: 38  
 Percent Similarity: 61.18% Conservative: 14  
 Best Local Similarity: 44.71% Mismatches: 32  
 Query Match: 41.47% Indels: 1  
 DB: 6 Gaps: 1

US-09-917-376-5 (1-88) x AAD26526 (1-6416)

QY 4 GlyVallysValGlnTyrIysAsnAspSerAlaProGlyAspAsnGlnIleIysPro 23  
 Db 2665 GGTGTGAAGTACTGTACAAACAATGAGCAAGTGCAGCACAGGTTCTATAAGCGC 2724  
 QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43  
 Db 2725 TGGTTTAAGTAGTAAGTGGAGGCAGCAGTGTGATCTTAGCAGGTTAAGATAAGA 2784  
 QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaLa 63  
 Db 2785 TACTGTACACAGTGGATGGTGACAGCCACAGAGTGGCGTA--TGTGACTGGGCACAG 2841  
 QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83  
 Db 2842 ATAGGGCAAGCAATGTGACATTCAATTTGTGAAGCTTAGCAGCGGAGTGAGTGGAGCG 2901  
 QY 84 AspThrTyrLeuGln 88  
 Db 2902 GATTATTACCTGGAG 2916

RESULT 11  
 AAX55661  
 ID AAX55661 standard; DNA; 11706 BP.  
 XX  
 AC AAX55661;  
 XX  
 DT 30-JUL-1999 (first entry)  
 XX  
 DE DNA sequence encoding truncated cellulases Cel B4/5 and Cel B5.  
 XX  
 KW Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;  
 KW Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;  
 KW cotton-containing fabric; stonewashing; ss.  
 XX  
 OS Unidentified.  
 XX  
 PN EP921188-A2.  
 XX  
 PD 09-JUN-1999.  
 XX  
 PF 15-SEP-1998; 98EP-00810919.  
 XX  
 PR 19-SEP-1997; 97US-00932571.  
 XX  
 PA (CLRN ) CLARIANT FINANCE BVI LTD.  
 XX  
 PI Anderson P, Bergquist PL, Daniels RM, Farrington GK, Gibbs MD;  
 PI Morgan H, Williams DP;  
 DR WPI: 1999-315403/27.  
 DR P-PSDB; AAY13492.





Db	4215	ATAGGGCAGCAATGTGACATTCAATTTTGTGGAAGCTGAGACGCGAGTGAGTGAGCG	4274
Qy	84	AspThrTyrLeuGln 88	
Db	4275	GATTATTACTTGGAG 4289	
RESULT 13			
AAL41028			
ID	AAL41028	standard; DNA; 1434 BP.	
XX	AC	AAL41028;	
XX	DT	11-OCT-2002 (first entry)	
XX	DE	CMCase gene with additional His encoding sequences.	
XX	KW	Protein surface-displayed; genetic carrier; spore; monoclonal variant;	
XX	KW	high-throughput screening; antibody; vaccine; enzyme; gene; ds.	
XX	OS	Unidentified.	
XX	XX		
XX	XX	Key Location/Qualifiers	
XX	FT	CDS 1..1434	
XX	FT	/*tag= a	
XX	FT	/product= "CMCase protein"	
XX	XX	WO200255561-A1.	
XX	XX	18-JUL-2002.	
XX	PD		
XX	PF	15-JAN-2002; 2002WO-KR000059.	
XX	XX	15-JAN-2001; 2001KR-00002156.	
XX	XX	(GENO-) GENOFOCUS CO LTD.	
XX	XX	Pan JG, Choi SK, Jung HC;	
XX	XX	WPI; 2002-590661/63.	
XX	XX	P-PSDB; AAO22444.	
XX	XX		
XX	PT	Methods for surface display of proteins, e.g. displaying a protein on the	
XX	PT	surface of spores, improving a protein or isolating a substance, by	
XX	PT	transforming a host cell harboring a genetic carrier with spore or virus	
XX	PT	with the vector library.	
XX	XX		
XX	PS	Example 5; Page 109-113; 118pp; English.	
XX	CC	The invention relates to methods for preparing a protein surface-	
XX	CC	displayed on a genetic carrier, improving a protein or isolating a	
XX	CC	substance in a mixture comprising transforming a host cell harbouring a	
XX	CC	genetic carrier consisting of a spore or virus with the vector library.	
XX	CC	The method is useful for surface display of proteins, particularly for	
XX	CC	displaying a protein on the surface of e.g. spores, for improving a	
XX	CC	protein, or for isolating a substance. These methods are useful in	
XX	CC	obtaining monoclonal variants from a large library, or in high-throughput	
XX	CC	screening of antibodies for use in therapy (e.g. as vaccines), diagnosis	
XX	CC	or analysis. This polynucleotide sequence represents the CMCase gene with	
XX	CC	additional His encoding sequences relating to the invention	
XX	SQ	Sequence 1434 BP; 493 A; 293 C; 341 G; 307 T; 0 U; 0 Other;	
Alignment Scores:			
Pred. No.:	1..94e-13	Length:	1434
Score:	186.00	Matches:	34
Percent Similarity:	61.18%	Conservative:	18
Best Local Similarity:	40.00%	Mismatches:	31
Query Match:	40.09%	Indels:	2
DB:	6	Gaps:	1
US-09-917-376-5 (1-88) x AAL41028 (1-1434)			
Qy	4	GlyValIysValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGlnIleLysPro	23

```

XX SQ Sequence 1488 BP; 501 A; 297 C; 357 G; 333 T; 0 U; 0 Other;
XX
XX Alignment Scores:
XX Pred. No.: 2,04e-13 Length: 1488
XX Score: 188.00 Matches: 34
XX Percent Similarity: 61.18% Conservative: 18
XX Best Local Similarity: 40.00% Mismatches: 31
XX Query Match: 40.09% Indels: 2
XX DB: 6 Gaps: 1
XX
XX US-09-917-376-5 (1-88) x AAL41025 (1-1488)
XX
QY 4 GlyValLysValGlnTyrIlyAsnAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
DB 1048 GGTATTTCGTACAATACAGACAGCGGGATGGAGTATGACACAGCAACCAATCCGTCGG 1107
QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
DB 1108 CAGCTTCAATAAATAAATAACGGCATACCGGTTGATTAAAGATGTCACGTCGCCGT 1167
QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63
DB 1168 TACTGGTATACCGCAAAACAAAGCCCAAAAGCTT-----GACTGTGACTACGGCAG 1221
QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
DB 1222 CTTGGATGCGCAATGTGACATACAAAGTTTGTGACGTTCGATTAACCAACCAAGGTGCA 1281
QY 84 AspThrTyrLeuGln 88
DB 1282 GATACCTATCTGGAA 1296
XX
XX RESULT 15
XX ID ABK53202
XX AC ABK53202;
XX DT 12-AUG-2002 (first entry)
XX DE Bacillus subtilis spore coat protein cotG-CMCase fusion gene.
XX
XX KW Spore coat protein; spore surface; delivery system; bioconversion;
XX KW recombinant live vaccine; resistance property; protein microarray;
XX KW whole cell absorber; secretion signal; target signal; cell membrane;
XX KW surface display motif; cotG; CMCase; carboxymethyl cellulase; gene; ds.
XX
XX OS Bacillus subtilis.
XX OS Synthetic.
XX OS Chimeric.
XX
XX FH Key
XX FT Promoter
XX FT 1..460
XX FT /tag= a
XX FT /note= "CotG promoter"
XX FT CDS
XX FT 461..2494
XX FT /tag= b
XX FT /partial
XX FT /product= "Bacillus subtilis spore coat protein cotG-
XX FT carboxymethyl cellulase (CMCase) fusion protein"
XX FT /note= "This sequence lacks a start codon"
XX FT misc_feature
XX FT 461..1045
XX FT /tag= c
XX FT /note= "CotG structural gene"
XX FT misc_feature
XX FT 1046..1084
XX FT /tag= d
XX FT /note= "Linker"
XX FT misc_feature
XX FT 1085..2491
XX FT /tag= e
XX FT /note= "CMCase structural gene"
XX
XX W0200246388-A1.

```

```

XX PD 13-JUN-2002.
XX
XX 07-DEC-2001; 2001WO-KR002124.
XX
XX 08-DEC-2000; 2000KR-00074835.
XX
XX (GENO-) GENOFOCUS CO LTD.
XX
XX PI Pan J, Choi S, Jung H;
XX
XX WPI: 2002-463830/49.
XX P-PSDB; AAU98063.
XX
XX Displaying a protein of interest on spore surface, by transforming a host
XX cell with a vector comprising a gene encoding a spore coat protein and a
XX gene encoding a protein of interest, useful for preparing protein
XX microarray.
XX
XX Example 4; Page 145; 158pp; English.
XX
XX The present invention relates to a new method for displaying a protein of
XX interest on a spore surface. The method of the invention involves
XX preparing a vector comprising a gene construct containing a gene encoding
XX a spore coat protein and a gene encoding a protein of interest,
XX transforming a host cell with the vector, displaying the protein of
XX interest on a surface of a spore of the host cell, and recovering the
XX spore. The method is useful for producing a delivery system for
XX recombinant live vaccine. The method is also useful for improving a
XX protein of interest using a resistance property of spore, in
XX bioconversion reactions, for preparing protein microarray, for producing
XX an antibody to antigen in vertebrates, and for preparing a whole cell
XX absorber. The spore coat proteins used circumvent a necessity for passage
XX across cell membrane, so that they do not need secretion signal and
XX target signal which are prerequisites of surface display motif, therefore,
XX ensuring a surface display of protein, such as beta-galactosidase, in
XX orderly fashion, which is difficult to pass across cell membrane. The
XX present nucleic acid sequence encodes the Bacillus subtilis spore coat
XX protein cotG-carboxymethyl cellulase (CMCase) fusion protein of the
XX invention
XX
XX SQ Sequence 2510 BP; 869 A; 528 C; 521 G; 592 T; 0 U; 0 Other;
XX
XX Alignment Scores:
XX Pred. No.: 3.97e-13 Length: 2510
XX Score: 188.00 Matches: 34
XX Percent Similarity: 61.18% Conservative: 18
XX Best Local Similarity: 40.00% Mismatches: 31
XX Query Match: 40.09% Indels: 2
XX DB: 6 Gaps: 1
XX
XX US-09-917-376-5 (1-88) x ABK53202 (1-2510)
XX
QY 4 GlyValLysValGlnTyrIlyAsnAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
DB 2051 GGTATTTCGTACAATACAGACAGCGGGATGGAGTATGACACAGCAACCAATCCGTCGG 2110
QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
DB 2111 CAGCTTCAATAAATAAATAACGGCAATACCGGTTGATTAAAGATGTCACGTCGCCGT 2170
QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63
DB 2171 TACTGGTATACCGCAAAACAAAGCCCAAAAGCTT-----GACTGTGACTACGGCAG 2224
QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
DB 2225 CTTGGATGCGCAATGTGACATACAAAGTTTGTGACGTTCGATTAACCAACCAAGGTGCA 2284
QY 84 AspThrTyrLeuGln 88
DB 2285 GATACCTATCTGGAA 2299

```

Search completed: May 12, 2004, 05:39:39  
Job time : 137.068 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 12, 2004, 04:33:52 ; Search time 826.547 Seconds  
(without alignments)  
3179.336 Million cell updates/sec

Title: US-09-917-376-5  
Perfect score: 489  
Sequence: 1 VSGGVKQYKNDSPAGDNQ.....IRASFGSVNPTATDYTLQ 88

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n\_model -DEV=xlh  
-Q=/cgm2\_1/USPRO\_spool/US09917376/runat\_11052004\_114535\_28249/app\_query.fasta\_1.1429  
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09917376 @CGN 1 5334 @runat\_11052004\_114535\_28249 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1.\*

29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	81	17.3	584	14	CF480372	CF480372 POL1_65 H
2	79.5	17.0	754	12	BQ047524	BQ047524 EST56642
3	78.5	16.7	768	14	CB293085	CB293085 UCRCS01_0
4	77	16.4	906	13	BQ722991	BQ722991 AGENCOURT
5	76	16.2	376	9	AA043716	AA043716 zk61f10.8
6	76	16.2	606	10	BE362891	BE362891 DGI_90 C1
7	76	16.2	641	14	CD226529	CD226529 CCL1_46 E
8	76	16.2	657	10	BE362820	BE362820 DGI_89 C0
9	76	16.2	676	14	CF487649	CF487649 POL1_45 C
10	76	16.2	691	12	BG739392	BG739392 EM1_82 D0
11	76	16.2	724	28	BZ616323	BZ616323 I958e08.9
12	76	16.2	814	12	B1218457	B1218457 602937951
13	76	16.2	905	28	BZ545807	BZ545807 CGALR72TM
14	76	16.2	934	29	CG270741	CG270741 OGWB33TH
15	76	16.2	1001	29	CG270748	CG270748 OGWB33TV
16	75.5	16.1	657	10	BF513128	BF513128 UI-H-BW1
17	75.5	16.1	1058	14	CK231562	CK231562 ILLUMIGEN
18	74.5	15.9	465	10	AW491464	AW491464 UI-M-BH3
19	74.5	15.9	613	14	CA595709	CA595709 wpalc.pk0
20	74.5	15.9	650	28	BZ894843	BZ894843 Hg4_0131
21	74	15.8	302	14	CF076729	CF076729 QHK12L22.
22	74	15.8	653	12	B1664037	B1664037 603289338
23	73.5	15.7	261	10	BF875908	BF875908 RC3-ET013
24	73.5	15.7	816	14	CB293980	CB293980 UCRCS01_0
25	73.5	15.7	819	14	CB293084	CB293084 UCRCS01_0
26	73.5	15.7	1039	12	BG334904	BG334904 607461219
27	73.5	15.7	1237	13	BQ065850	BQ065850 AGENCOURT
28	73	15.6	357	14	CF078232	CF078232 QHK1J23.Y
29	73	15.6	360	13	C64738	C64738 C64738 YuJ1
30	73	15.6	593	14	CF622108	CF622108 laf13d06.
31	73	15.6	776	29	BK209426	BK209426 Danio rer
32	73	15.6	861	28	BH319803	BH319803 CH230-115
33	73	15.6	894	29	CG077853	CG077853 PUFKJ89TB
34	73	15.6	924	29	CG077856	CG077856 PUFKJ89TD
35	72.5	15.5	380	12	BG816375	BG816375 dad35b08.
36	72.5	15.5	450	9	AL799805	AL799805 AL799805
37	72.5	15.5	533	10	AW963120	AW963120 EST375193
38	72.5	15.5	544	9	AL898643	AL898643 AL898643
39	72.5	15.5	563	28	BZ424716	BZ424716 100018538
40	72.5	15.5	567	14	CF754312	CF754312 EST-77-2-
41	72.5	15.5	579	9	AL962576	AL962576 AL962576
42	72.5	15.5	601	29	CE253602	CE253602 tigr-gss-
43	72.5	15.5	640	9	AL846493	AL846493 AL846493
44	72.5	15.5	648	12	BM609423	BM609423 170006871
45	72.5	15.5	719	12	BG635313	BG635313 AT31857.5

ALIGNMENTS

RESULT 1  
CF480372  
LOCUS  
DEFINITION  
POL1\_65\_H07.b1\_A002 Pollen Sorghum bicolor CDNA clone  
ACCESSION  
CF480372  
VERSION  
CF480372.1 GI:34509241  
KEYWORDS  
EST  
SOURCE  
Sorghum bicolor (sorghum)  
ORGANISM  
Sorghum bicolor  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE	1 (bases 1 to 584)	Db	213	GCTTGTGTGAAGTGTGGTGTGCTTGTGAGACGGTGTCTGTCTGTGTGTCG 272
AUTHORS	Cordonnier-Pratt,M.-M., Suzuki,Y., Sugano,S., Klein,R.R., Liang,C., Sun,F., Sullivan,R., Eastman,A., Cannon,R., Kern,B., Morgan,J., Lucas,A., Al-Sheikh,A., Jones,V., Adibi,N., Owen,A., Gao,J. and Pratt,L.H.	Qy	79	AlaThrProThrAlaAsp 84
TITLE	EST database from Sorghum: pollen	Db	273	AGCTTCAGTCTGTGTGAT 290
JOURNAL	Unpublished (2003)	RESULT 2		
COMMENT	Other ESTs: POL1.65.H07.g1.A002 Contact: Cordonnier-Pratt MM Laboratory for Genomics and Bioinformatics The University of Georgia, Department of Plant Biology Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA Tel: 706 542 1860 Fax: 706 583 0210 Email: mmpratt@uga.edu Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in the Human Genome Center, University of Tokyo Institute of Medical Science; plant material and RNA prepared at Texas A & M University; sequencing done in the Laboratory for Genomics and Bioinformatics, University of Georgia. Sequence ends have been trimmed to exclude vector and regions below Phred quality 16. Three-prime sequences are presented as their reverse complement and have been trimmed to exclude polyA. Seq primer: Sug3-14 (TAGTCTAGCGCGCGGACC) POLYA=yes.	LOCUS	BQ047524	754 bp mRNA linear EST 10-MAR-2003
		DEFINITION		EST596642 P. infestans-challenged potato leaf, incompatible reaction Solanum tuberosum cDNA clone BPL117P18 5' end, mRNA sequence.
		ACCESSION		BQ047524
		VERSION		BQ047524.1 GI:19821510
		KEYWORDS		EST.
		SOURCE		Solanum tuberosum (potato)
		ORGANISM		Solanum tuberosum
		REFERENCE		1 (bases 1 to 754)
		AUTHORS		Zhang,P., Hernandez,M., Tornqvist,C.-E., Wirtz,U., Loukoianov,A., Rangel,P., Haberland,G.T., Karanycheva,S.A., Tsai,J., Chiemiango,A., Bougri,O., Buell,C.R., Ronning,C.M., Helgeson,J. and Baker,B.
		TITLE		Generation of ESTs from Potato Leaves Challenged with Phytophthora infestans, incompatible interaction (2002)
		JOURNAL		Unpublished (2002)
		COMMENT		Contact: Robin Buell The Institute for Genomic Research 9712 Medical Center Dr, Rockville, MD 20850, USA Email: potato-array@tigr.org This clone can be obtained from the University of Arizona Genomics Institute. Orders can be made through URL: http://genome.arizona.edu/orders/ Seq primer: T3.
FEATURES	Location/Qualifiers	FEATURES		Location/Qualifiers
source	1..584	source	1..754	
	/organism="Sorghum bicolor"		/organism="Solanum tuberosum"	
	/mol_type="mRNA"		/mol_type="mRNA"	
	/cultivar="BTx623"		/cultivar="Kennebec"	
	/db_xref="taxon:4558"		/db_xref="taxon:4113"	
	/clone="POL1.65.H07.A002"		/clone="BPL117P18"	
	/lab_host="DH10B-T1 phage-resistant E. coli"		/tissue_type="leaf"	
	/clone_lib="Pollen"		/dev_stage="6 week old"	
	/note="Organ: Pollen; Vector: pME18S-FL3; Site 1: XhoI; Site 2: XhoI; The library was prepared from polyA+ RNA from pollen at the late vacuolated-vacuolated stage of development. Pollen was harvested from greenhouse-grown panicles of sorghum line BTx623. Panicles were removed from the flag leaf prior to emergence, when no detectable anylase is present in pollen of male-fertile lines. This stage represents pollen collected from anthers about 8-14 days prior to anthesis. Double-stranded cDNA was cloned unidirectionally into different DraIII sites of the pME18S-FL3 vector (5-prime DraIII site is CACTGTGTG, 3-prime DraIII site is CACCATGTG). XhoI excises the cDNA insert."		/lab_host="SOLR"	
			/clone_lib="P. infestans-challenged potato leaf, incompatible reaction"	
			/notes="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2: XhoI; supplier: UC Berkeley, PGEC; sequencing: The Institute for Genomic Research. Whole plants were challenged with 450,000 sporangia/ml P. infestans isolate US-1 (US940501) in Biotron (Madison, Wisconsin). Leaf tissue was collected at 1, 2, 5, 12, and 24 hours post-challenge and frozen in liquid nitrogen immediately upon removal. Kennebec plants showed no signs of HR. Katahdin plants (susceptible to P. infestans US-1) were used as controls and showed infection. NOTE: We cannot exclude the possibility that this sequence is actually derived from Phytophthora rather than potato."	
ORIGIN		ORIGIN		
Alignment Scores:		Alignment Scores:		
Pred. No.:	12.7	Pred. No.:	27.9	Length: 754
Score:	81.00	Score:	79.50	Matches: 29
Percent Similarity:	45.35%	Percent Similarity:	38.05%	Conservative: 14
Best Local Similarity:	27.91%	Best Local Similarity:	25.66%	Mismatches: 33
Query Match:	17.27%	Query Match:	16.95%	Indels: 37
DB:	14	DB:	12	Gaps: 4
US-09-917-376-5 (1-88) x CF480372 (1-584)		US-09-917-376-5 (1-88) x BQ047524 (1-754)		
Qy	5	Qy	7	ValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGlnIleLysProGly 24
Db	45	Db	45	ATAGCAATCAGCGAAGCACTCAATGAATCGGTCCGGAGGAGCCCTGTGTCGCCGGGC 104
Qy	25	Qy	25	LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg--- 43
Db	105	Db	105	-----CGCGCCGGGACAGAACCTGAAACTTCCAAACGATCGTGCAGGAAA 152
Qy	44	Qy	44	---TyrTrp---PheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61
Db	153	Db	153	CCATGTGGGGATCTGCGGGAAGAGGTTCAGCACCATGGAATATATATATGATGAA 212
Qy	62	Qy	62	AlaAlaMetGlyCysGlyAsnIle-----ArgAlaSerPheGlySerValAsnPro 78

Db 152 GTGCTTATCAAGTCCAGGAAGT---CCGGTAGTAACTAATATCTCCTGGTTCAGTA 208  
 Qy 27 LeuValAenThrGlySerSerSer----- 34  
 Db 209 GTATCAAAATTCGGCACCTCCTCACTTCCCTGGAAAATGCCCTATATTGAGTTTCGT 268  
 Qy 35 -----ValAspLeuSerThrValThrValArgTyrTrpPheThr 47  
 Db 269 AAGGGGGAGCCTCCGAAGTCTTCTGGTTATGAACATTTCTCCACTCGCAAAATGGGGTTCA 328  
 Qy 48 ArgAspGly-----GlySerSerThr 54  
 Db 329 AGGGTTGGCTCAGGATCATTCACACCAAGTGGCTGGGCTCTAGCTAGTTCTGGAAC 388  
 Qy 55 LeuValTyrAenCysAspTrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGly 74  
 Db 389 CTGACCCCAATGTTGGGATTTCAAGGCTAGGTTCTGGTACTGTGCTCACTCCAAATGGTGG 448  
 Qy 75 SerValAenProAlaThrProThrAlaAspThrTyrLeu 87  
 Db 449 GAA-----CCTCCTCCCGAGATAGTTACCTT 475

RESULT 3  
 LOCUS CB293085 768 bp mRNA linear EST 28-FEB-2003  
 DEFINITION UCRCS01\_05cb12\_g1 Washington Navel orange cold acclimated flavedo & albedo cDNA library Citrus sinensis cDNA clone UCRCS01\_05cb12, mRNA sequence.  
 ACCESSION CB293085  
 VERSION CB293085.1 GI:28618542  
 LOCUS EST.  
 KEYWORDS Citrus sinensis  
 ORGANISM Citrus sinensis  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Sapindales; Rutaceae; Citrus.  
 REFERENCE 1 (bases 1 to 768)  
 AUTHORS Close,T.J., Collura,K., Fenton,R.D., Feuerbacher,O., Kim,H.R., Kudrna,D., Wananaker,S., Wing,R. and Yu,Y.  
 TITLE Development of EST Resources and New Genetic Markers for California Citrus  
 JOURNAL Unpublished (2003)  
 COMMENT Contact: Timothy Close  
 Department of Botany & Plant Sciences, University of California  
 Riverside, CA, 92521-0124  
 Tel: 909/7873318  
 Fax: 909/7874437  
 Email: timothy.close@ucr.edu  
 Seq primer: T3.

FEATURES  
 source  
 1..768  
 /organism="Citrus sinensis"  
 /mol\_type="mRNA"  
 /cultivar="Washington navel"  
 /db\_xref="taxon:2711"  
 /clone="UCRCS01\_05cb12"  
 /tissue\_type="rind containing flavedo and albedo"  
 /dev\_stage="Mature fruit"  
 /lab\_host="E. coli TJC121"  
 /clone\_lib="Washington Navel orange cold acclimated flavedo & albedo cDNA library"  
 /note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site 1: EcoRI; Site 2: XhoI; Plants were grown in the field at University of California, Riverside Agricultural Operations since 1983. The scion was Washington Navel orange and the rootstock Carizzo Citrange. Tissue from mature fruit was harvested at mid-day in January 2002 during a cold spell, when pre-dawn temperatures were approximately -2 to 2 degree C. Approximately 2 cm median sections of the rind were excised in the field from several fruits, then wrapped in aluminum foil and frozen quickly in dry ice. Total RNA was extracted using a phenol extraction procedure described in J. Japanese Soc. Hort.

Sci. (1996) 64:809-814. Poly(A) RNA was purified, a cDNA library was made, and 1 million primary lambda cDNA clones were in vivo excised to give a population of phuscript SK(-) phagemids. All steps to this point were performed in the TJ Close lab at the University of California, Riverside (Fenton). Phagemids were plated, plasmid DNA purified, cDNA clones archived, and DNA sequences determined bi-directionally using an ABI3530 at the Arizona Genomics Institute, University of Arizona (Collura, Feuerbacher, Kim, Kudrna, Wing, Yu). Chromatogram files were transmitted to UC Riverside (by Yu), then processed at UC Riverside (by Wananaker) using the Harvest pipeline (<http://harvest.ucr.edu>) to remove vector and cloning oligo sequences and various contaminants, and to trim to a high quality region. Sequences that retained a phred 17 region of at least 100 bases were deposited to GenBank."

## ORIGIN

Alignment Scores:  
 Pred. No.: 37.9 Length: 768  
 Score: 78.50 Matches: 20  
 Percent Similarity: 57.69% Conservative: 10  
 Best Local Similarity: 38.46% Mismatches: 15  
 Query Match: 16.74% Indels: 7  
 DB: 14 Gaps: 2

US-09-917-376-5 (1-88) x CB293085 (1-768)

Qy 3 GlyGlyValIysValGlnTyrIysAsnAsnAspSerAlaProGlyAspAsnGlnIleIys 22  
 Db 612 GCGCGCATCAAC-----AAGAACACACACAGCAGCAGCAATAATAATAAGTAAT 662  
 Qy 23 ProGlyLeuGlnLeuValAenThrGlySerSerSerValAspLeuSerThrValThrVal 42  
 Db 663 AATAATAGCGACGTTTACTAGTGGTAGCGTAGCGTGTGTCGAAGTCCGTGACG--- 719  
 Qy 43 ArgTyrTrpPheThrArgAspGlyGlySerSerThr 54  
 Db 720 -----TTTTCGGAGCGGAGGAGGAGCAGCAGT 746

## RESULT 4

LOCUS BQ722991 906 bp mRNA linear EST 16-JUL-2002  
 DEFINITION AGENCOURT 8103733 Lupeki sympathetic\_trunk Homo sapiens cDNA clone IMAGE:6191052 5', mRNA sequence.

ACCESSION BQ722991  
 VERSION BQ722991.1 GI:21861888  
 KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 906)  
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)

Tissue Procurement: Dr. James R. Lupski

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM13590 row: n column: 13

High quality sequence stop: 512.

Location/Qualifiers

## FEATURES

## source

1..906

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"



Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 583 0210  
 Email: mmpratt@uga.edu  
 Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20. Three-prime sequences, which are obtained with PolyTMix or T7 sequencing primer, are presented as the reverse complement.  
 Seq primer: PolyTMix  
 High quality sequence start: 15  
 High quality sequence stop: 594  
 POLYA=No.

#### FEATURES

source  
 1..606  
 /organism="Sorghum bicolor"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:4558"  
 /clone\_lib="Dark Grown 1 (DGI)"  
 /note="Organ: 5-day-old dark-grown seedlings; Vector: Lambda Zap; Site 1: XhoI; Site 2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."

#### ORIGIN

Alignment Scores:  
 Pred. No.: 55.4 Length: 606  
 Score: 76.00 Matches: 23  
 Percent Similarity: 44.19% Conservative: 15  
 Best Local Similarity: 26.74% Mismatches: 38  
 Query Match: 16.20% Indels: 10  
 DB: 10 Gaps: 4

US-09-917-376-5 (1-88) x BE362891 (1-606)

QY 5 ValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnLeLysProGly 24  
 DB 208 ATAGCAATCAGCGGAACTCAATGAATCCGTCGGGAGGAGCGCTGTCTGCGCGGCG 267  
 QY 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg--- 43  
 DB 268 -----CGCGCGGGGACAGAACCTTCCAAACGATCGGTGCAAGGAAA 315  
 QY 44 ---TyrTrp---PheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61  
 DB 316 CCATGGTGGGATCTCGAGGAGGAGGAGTTTGACCAACATGATATATATATGATGAA 375  
 QY 62 AlaAlaMetGlyCysGlyAsnLe-----ArgAlaSerPheGlySerValAsnPro 78  
 DB 376 GCTTGTGTGAAGTGTGGTGTGTTCTTTTGGAGACGGTGTCTTGTCTGTGTGTCG 435  
 QY 79 AlaThrProThrAlaAsp 84  
 DB 436 AGCTTGCAGTCTGTGTAT 453

RESULT 7  
 CD226529 641 bp mRNA linear EST 21-MAY-2003  
 LOCUS CCC1\_46\_E07\_b1\_A007 Callus culture/cell suspension Sorghum bicolor  
 DEFINITION cDNA clone CCC1\_46\_E07\_A007 3', mRNA sequence.

ACCESSION CD226529  
 VERSION CD226529.1 GI:30969963  
 KEYWORDS EST.

Sorghum bicolor (sorghum)  
 Sorghum bicolor  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE 1 (bases 1 to 641)  
 Cordonnier-Pratt M.-M., Wentzel, V., Suzuki, Y., Sugano, S.,  
 Klein, R.R., Liang, C., Sun, F., Sullivan, R., Shah, M., Rathore, K.,  
 Eastman, A. and Pratt, L.H.  
 An EST database from Sorghum: callus culture and cell suspension  
 Unpublished (2003)

#### COMMENT

Other ESTs: CCC1\_46\_E07\_g1\_A007  
 Contact: Cordonnier-Pratt MM  
 Laboratory for Genomics and Bioinformatics  
 The University of Georgia, Department of Plant Biology  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 583 0210  
 Email: mmpratt@uga.edu  
 Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in  
 the Human Genome Center, University of Tokyo Institute of Medical  
 Science; plant material and RNA prepared at Texas A & M University;  
 sequencing done in the Laboratory for Genomics and Bioinformatics,  
 University of Georgia. Sequence ends have been trimmed to exclude  
 vector and regions below Phred quality 16. Three-prime sequences  
 are presented as their reverse complement and have been trimmed to  
 exclude polyA.  
 Seq primer: Sug3 (CGACCTGCAGCTCGAGCACA)  
 POLYA=Yes.

#### FEATURES

source  
 1..641  
 /organism="Sorghum bicolor"  
 /mol\_type="mRNA"  
 /cultivar="RTx430"  
 /db\_xref="taxon:4558"  
 /clone="CCC1\_46\_E07\_A007"  
 /lab\_host="DH10B-T1 phage-resistant E. coli"  
 /clone\_lib="Callus culture/cell suspension"  
 /note="Vector: pME18S-FL3; Site 1: XhoI; Site 2: XhoI; The library was prepared from a mixture of polyA+ RNA from callus culture tissue and cells in suspension culture. Double-stranded cDNA was cloned unidirectionally into different DraIII sites of the pME18S-FL3 vector (5-prime DraIII site is CACTGTGTG, 3-prime DraIII site is CACCATGTG). XhoI excises the cDNA insert."

#### ORIGIN

Alignment Scores:  
 Pred. No.: 59.9 Length: 641  
 Score: 76.00 Matches: 23  
 Percent Similarity: 44.19% Conservative: 15  
 Best Local Similarity: 26.74% Mismatches: 38  
 Query Match: 16.20% Indels: 10  
 DB: 14 Gaps: 4  
 US-09-917-376-5 (1-88) x CD226529 (1-641)

QY 5 ValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnLeLysProGly 24  
 DB 149 ATAGCAATCAGCGGAACTCAATGAATCCGTCGGGAGGAGCGCTGTCTGCGCGGCG 208  
 QY 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg--- 43  
 DB 209 -----CGCGCGGGGACAGAACCTTCCAAACGATCGGTGCAAGGAAA 256  
 QY 44 ---TyrTrp---PheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61  
 DB 257 CCATGGTGGGATCTCGAGGAGGAGGAGTTTGACCAACATGATATATATATGATGAA 316  
 QY 62 AlaAlaMetGlyCysGlyAsnLe-----ArgAlaSerPheGlySerValAsnPro 78  
 DB 317 GCTTGTGTGAAGTGTGGTGTGTTCTTTGAGACGGTGTCTTGTCTGTGTGTCG 376  
 QY 79 AlaThrProThrAlaAsp 84  
 DB 377 AGCTTGCAGTCTGTGTAT 394

#### RESULT 8

BE362820 657 bp mRNA linear EST 20-JUL-2000  
 LOCUS DGI\_89\_C02\_g2\_A002 Dark Grown 1 (DGI) Sorghum bicolor cDNA, mRNA  
 DEFINITION sequence.  
 ACCESSION BE362820  
 VERSION BE362820.1 GI:9304377



```

KEYWORDS EST.
SOURCE Sorghum bicolor (sorghum)
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 657)
AUTHORS Cordonnier-Pratt, M.-M., Gingle, A., Marsala, C., Sudman, M. and
Pratt, L.H.
TITLE An EST database from Sorghum: dark-grown seedlings
JOURNAL Unpublished (2000)
COMMENT Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: PolyTWix
High quality sequence start: 89
High quality sequence stop: 651
POLYA=NO.

FEATURES
source
1..657
Location/Qualifiers
/organism="Sorghum bicolor"
/mol_type="mRNA"
/db_xref="taxon:4558"
/clone_lib="Dark Grown 1 (DG1)"
/note="Organ: 5-day-old dark-grown seedlings; Vector:
Lambda Zap; Site 1: XhoI; Site 2: EcoRI; The library was
made from poly-A RNA in the cloning vector lambda Zap II.
Clones to be sequenced were prepared by mass excision."

ORIGIN
Alignment Scores: 62 Length: 657
Pred. No.: 76.00 Matches: 23
Score: 44.19% Conservative: 15
Percent Similarity: 26.74% Mismatches: 38
Best Local Similarity: 16.20% Indels: 10
Query Match: 10 Gaps: 4
DB: 10

US-09-917-376-5 (1-88) x BE362820 (1-657)

QY 5 ValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnLeLysProGly 24
::: ::::: |||||
Db 204 ATAGCAATCAGCGCGAACTCAATGAATCCGGTCCGGAGGAGCCTGTGCTGGCCGGGC 263
::: ::::: |||||
QY 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg--- 43
::: ::::: |||||
Db 264 -----CGCGCCGGGAGAGAACCCCTGAAACTTCCAAACGATCGGTGCAAGGAAA 311
::: ::::: |||||
QY 44 ---TyrTrp---PheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61
::: ::::: |||||
Db 312 CAGTGTGGGGATCTGCGAGGAGGAGGTTGAGCAACATGAATATATATATATGATGAA 371
::: ::::: |||||
QY 62 AlaAlaMetGlyCysGlyAsnLeu-----ArgAlaSerPheGlySerValAsnPro 78
||| ||| ||||| |||||
Db 372 GCTGTGTGAAGTGTGGTACGTTGTTCTTTGAGACGGTGTCTGCTCTGTGTGTCG 431
::: ::::: |||||
QY 79 AlaThrProThrAlaAsp 84
::: ::::: |||||
Db 432 AGCTTCAGTCTGGTGAT 449
::: ::::: |||||

RESULT 9
CF487649 676 bp mRNA linear EST 08-SEP-2003
LOCUS POL1_45_C08_b1 A002 Pollen Sorghum bicolor cDNA clone
DEFINITION POL1_45_C08_A002 3', mRNA sequence.
ACCESSION CF487649

```

VERSION  
KEYWORDS  
SOURCE  
ORGANISM

CF487649.1 GI:34516518  
EST.  
Sorghum bicolor (sorghum)  
Sorghum bicolor

REFERENCE  
AUTHORS

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Sorghum.  
1 (bases 1 to 676)  
Cordonnier-Pratt, M.-M., Suzuki, Y., Cannon, S., Klein, R.R., Liang, C.,  
Sun, F., Sullivan, R., Eastman, A., Cannon, R., Kern, B., Morgan, J.,  
Lucas, A., Al-Sheikh, A., Jones, V., Adibi, N., Owen, A., Gao, J. and  
Pratt, L.H.

TITLE  
JOURNAL  
COMMENT

EST database from Sorghum: pollen  
Unpublished (2003)  
Other ESTs: POL1\_45\_C08\_g1-A002  
Contact: Cordonnier-Pratt MM  
Laboratory for Genomics and Bioinformatics  
The University of Georgia, Department of Plant Biology  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 583 0210  
Email: mmpratt@uga.edu

FEATURES  
source

Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in  
the Human Genome Center, University of Tokyo Institute of Medical  
Science; plant material and RNA prepared at Texas A & M University;  
sequencing done in the Laboratory for Genomics and Bioinformatics,  
University of Georgia. Sequence ends have been trimmed to exclude  
vector and regions below Phred quality 16. Three-prime sequences  
are presented as their reverse complement and have been trimmed to  
exclude polyA.  
Seq primer: Sug3-14 (TAGCTAGCGCGCGGACGACC)  
POLYA=Yes.

Location/Qualifiers

1..676  
/organism="Sorghum bicolor"  
/mol\_type="mRNA"  
/cultivar="BTx623"  
/db\_xref="taxon:4558"  
/clone="POL1\_45\_C08\_A002"  
/lab\_host="DH10B-T1 phage-resistant E. coli"  
/clone\_lib="Pollen"

ORIGIN

Alignment Scores: 64.6 Length: 676  
Pred. No.: 76.00 Matches: 23  
Score: 44.19% Conservative: 15  
Percent Similarity: 26.74% Mismatches: 38  
Best Local Similarity: 16.20% Indels: 10  
Query Match: 14 Gaps: 4  
DB: 14

US-09-917-376-5 (1-88) x CF487649 (1-676)

QY 5 ValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnLeLysProGly 24  
::: ::::: |||||  
Db 214 ATAGCAATCAGCGCGAACTCAATGAATCCGGTCCGGAGGAGGAGCCTGTGCTGGCCGGGC 273  
::: ::::: |||||  
QY 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg--- 43  
::: ::::: |||||  
Db 274 -----CGCGCCGGGAGAGAACCCCTGAACTTCCAAACGATCGGTGCAAGGAAA 321  
::: ::::: |||||

ORIGIN

Alignment Scores: 64.6 Length: 676  
Pred. No.: 76.00 Matches: 23  
Score: 44.19% Conservative: 15  
Percent Similarity: 26.74% Mismatches: 38  
Best Local Similarity: 16.20% Indels: 10  
Query Match: 14 Gaps: 4  
DB: 14

US-09-917-376-5 (1-88) x CF487649 (1-676)

QY 5 ValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnLeLysProGly 24  
::: ::::: |||||  
Db 214 ATAGCAATCAGCGCGAACTCAATGAATCCGGTCCGGAGGAGGAGCCTGTGCTGGCCGGGC 273  
::: ::::: |||||  
QY 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg--- 43  
::: ::::: |||||  
Db 274 -----CGCGCCGGGAGAGAACCCCTGAACTTCCAAACGATCGGTGCAAGGAAA 321  
::: ::::: |||||

ORIGIN

Alignment Scores: 64.6 Length: 676  
Pred. No.: 76.00 Matches: 23  
Score: 44.19% Conservative: 15  
Percent Similarity: 26.74% Mismatches: 38  
Best Local Similarity: 16.20% Indels: 10  
Query Match: 14 Gaps: 4  
DB: 14

US-09-917-376-5 (1-88) x CF487649 (1-676)

QY 5 ValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnLeLysProGly 24  
::: ::::: |||||  
Db 214 ATAGCAATCAGCGCGAACTCAATGAATCCGGTCCGGAGGAGGAGCCTGTGCTGGCCGGGC 273  
::: ::::: |||||  
QY 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg--- 43  
::: ::::: |||||  
Db 274 -----CGCGCCGGGAGAGAACCCCTGAACTTCCAAACGATCGGTGCAAGGAAA 321  
::: ::::: |||||

Qy 44 ---TyrTrp---PheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61  
 Db 322 CCATGTTGGGGGATCTGCAGGGAAGAGGTTTGACCAACATGGAATATATATATGAA 381  
 Qy 62 AlaAlaMetGlyCysGlyAsnIle-----ArgAlaSerPheGlySerValAsnPro 78  
 Db 382 GCTTGTGAAGTGTGGGACGTTGGTCTTCTTGGAGACGGTGTCTGTGTTGGTCCG 441  
 Qy 79 AlaThrProThrAlaAsp 84  
 Db 442 AGCTTGCACTCTGTGAT 459

RESULT 10  
 BG739392  
 LOCUS EMI\_82\_D05.g1\_A002 Embryo 1 (EM1) Sorghum bicolor cDNA, mRNA  
 DEFINITION BG739392  
 ACCESSION BG739392  
 VERSION BG739392.1 GI:14089081  
 KEYWORDS EST  
 SOURCE Sorghum bicolor (sorghum)  
 ORGANISM Sorghum bicolor  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Sorghum.  
 1 (bases 1 to 691)  
 Reid, S.P., Cordonnier-Pratt, M.-M., Gingle, A. and Pratt, L.H.  
 An EST database from Sorghum: developing embryos  
 Unpublished (2000)  
 CONTACT: Cordonnier-Pratt MM  
 LABORATORY for Genomics and Bioinformatics  
 The University of Georgia, Department of Plant Biology  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 583 0210  
 Email: mmpratt@uga.edu  
 Sequences have been trimmed to exclude polyA, vector and regions  
 below Phred quality 16. The threshold for highest quality sequence  
 is 20.  
 Seq primer: PolyTMix  
 High quality sequence start: 50  
 High quality sequence stop: 691  
 POLYA=No.

FEATURES  
 source  
 1..691  
 /organism="Sorghum bicolor"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:4558"  
 /clone\_lib="Embryo 1 (EM1)"  
 /note="Organ: Embryos germinated for 24 hr; Vector: pBluescript II from lambda Zap II; Site\_1: XhoI; Site\_2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda ZAP II. Clones to be sequenced were prepared by mass excision."  
 Alignment Scores:  
 Pred. No.: 66.6 Length: 691  
 Score: 76.00 Matches: 23  
 Percent Similarity: 44.19% Conservative: 15  
 Best Local Similarity: 26.74% Mismatches: 38  
 Query Match: 16.20% Indels: 10  
 DB: 12 Gaps: 4

US-09-917-376-5 (1-88) x BG739392 (1-691)  
 Qy 5 VallysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysProGly 24  
 Db 265 ATAGCAATCAGCCGAAATCAATCCGTCGCGGAGGAGCCCTGTGCTGCCGCGGC 324  
 Qy 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuThrValThrValArg--- 43

Db 325 -----CGCGCCGGGACAGAACCTTGAACATTCACACGATCGGTGCAAGGAA 372  
 Qy 44 ---TyrTrp---PheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61  
 Db 373 CCATGTTGGGGGATCTGCAGGGAAGAGGTTTGAGCAACATGGAATATATATGAA 432  
 Qy 62 AlaAlaMetGlyCysGlyAsnIle-----ArgAlaSerPheGlySerValAsnPro 78  
 Db 433 GCTTGTGAAGTGTGGTGAACGTTGGTCTTCTTGGAGACGGTGTCTGTGTTGGTCCG 492  
 Qy 79 AlaThrProThrAlaAsp 84  
 Db 493 AGCTTGCACTCTGTGAT 510

RESULT 11  
 BZ616323/c  
 LOCUS BZ616323  
 DEFINITION 724 bp DNA linear GSS 16-JAN-2003  
 ig58e08.g1 WGS-ZmaySF (DH5a methyl filtered) Zea mays genomic clone  
 ig58e08.5, genomic survey sequence.

ACCESSION BZ616323  
 VERSION BZ616323.1 GI:27766878  
 KEYWORDS GSS.  
 SOURCE Zea mays  
 ORGANISM Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Zea.  
 1 (bases 1 to 724)  
 Rabinowicz, P.D., O'Shaughnessy, A.L., Ballia, V., Dedhia, N.,  
 Katzenburger, F., King, L., Miller, B., Muller, S., Nascimento, L.,  
 Zutavern, T., McCombie, W.R. and Martienssen, R.A.  
 Genomic shotgun sequences from Zea mays (methyl-filtered)  
 Unpublished (2002)  
 CONTACT: W. Richard McCombie  
 Lita Annenberg Hazen Genome Sequencing Center  
 Cold Spring Harbor Laboratory  
 PO Box 100, Cold Spring Harbor, NY 11724, USA  
 Tel: 516 367 8884  
 Fax: 516 367 8874  
 Email: mcombie@cshl.org  
 Plate: ig58 row: e column: 08  
 Seq primer: -21M13UnivRev  
 Class: shotgun  
 High quality sequence stop: 728.

FEATURES  
 source  
 1..724  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /cultivar="B73"  
 /db\_xref="taxon:4577"  
 /clone="ig58e08"  
 /lab\_host="DH5a"  
 /clone\_lib="WGS-ZmaySF (DH5a methyl filtered)"  
 /note="Organ: immature ears; Site 1: Xba I; Site 2: Xba I;  
 The vector was digested with XbaI and one nucleotide was  
 added by fill in in the recessive 3' end. The genomic DNA  
 was nebulized, end repaired, adaptor ligated and size  
 fractionated using sephadex. The resulting fragments were  
 between 0.8 and 3 kb and were cloned into the vector (.x/y  
 reads in M13mp19, .b/g reads in pUC19). The same ligation  
 was transformed into DH5a."

ORIGIN  
 Alignment Scores:  
 Pred. No.: 71 Length: 724  
 Score: 76.00 Matches: 34  
 Percent Similarity: 36.28% Conservative: 7  
 Best Local Similarity: 30.09% Mismatches: 29  
 Query Match: 16.20% Indels: 43  
 DB: 28 Gaps: 6  
 US-09-917-376-5 (1-88) x BZ616323 (1-724)

```

QY 2 SerGlyGlyValLysValGlnTyrLysAsnAspSer-----AlaPro 16
DB TCCGAGCGGTGAGTATACCAATACCAAGGAGGAGCAAGCTCGCGAGCTCGGCC 528
QY 17 Gly-----AspAsnGlnLileLysProGlyLeuGlnLeuValAsn----- 29
DB GCGGCCAGCTGACGAGCTGCGGAGGAGCCACGACTGATCTTATCAACGTCGCCGC 468
QY 30 Thr-GlySerSerValAsp----- 36
DB ACGAGGAAGTTCGTCACTCCCGGAGATCGTCGGAGCGGTCAGAGCGCTGGTTCGAG 408
QY 37 -----LeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySerSe 53
DB GTCATCCGATCGAGCGCGGTAGACCTTCACGCTGGAGGACTTCGCCCGGACCGTCGAC 348
QY 53 rThrLeuValTyrAsnCyAspTTPAlaAlaMetGlyCyGlyAsnIleArgAlaSerPh 73
DB TCGTCCG-----ACGTGCTCATGCGCGGCGACGGG----- 317
QY 73 eGlySerValAsnProAlaThrProThrAlaAspThr 85
DB -----CGGCTCACCACACTTCTTCTCC 293

RESULT 12
BZ545807
LOCUS BZ545807 814 bp mRNA linear EST 11-JUL-2001
DEFINITION 602937951F1 NCI_CGAP_L19 Mus musculus cDNA clone IMAGE:5101069 5',
ACCESSION BZ545807
VERSION BZ545807.1 GI:14671901
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 814)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11243 row: f column: 14
High quality sequence stop: 792.
FEATURES             Location/Qualifiers
     source
     1..814
     /organism="Mus musculus"
     /mol_type="mRNA"
     /strain="FVB/N"
     /db_xref="taxon:10090"
     /clone="IMAGE:5101069"
     /lab_host="DH10B (T1 phage-resistant)"
     /clone_lib="NCI CGAP L19"
     /note="Organ: liver; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.9 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

ORIGIN
Alignment Scores:
Pred. No.:      83.6      Length:      814
Score:          76.00     Matches:     28
Percent Similarity: 44.83%  Conservative: 11
Best Local Similarity: 32.18% Mismatches: 29
Query Match:      16.20%  Indels:      19
DB:              12      Gaps:         4

US-09-917-376-5 (1-88) x BZ545807 (1-905)

```

```

US-09-917-376-5 (1-88) x BZ18457 (1-814)
QY 2 SerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21
DB GAGGTGGAGCCGAGTATCGGAGCGGAACACAGATCGCAAGGACAATGAGCAG--- 89
QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
DB -----AGTTCCCGAGTGGACCTCGAGGACCTG--- 116
QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeu-----ValTyrAsnCys 59
DB :::::::::::::::::::: 117
CTCCGCTACTTACAACAGAGCGCGCGGCTCTCACACGATCCACGCGGATGTACGGCTGT 176
QY 60 Asp-TpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAl 79
DB :::::::::::::::::::: 177
GACGTGGGTGGACGGCGCC-----TCCTCCGGGTACGAGCAGGTGCGCA 224
QY 79 aThrProThrAlaAspThr 85
DB TACGACGGCTCGGATTACA 243

RESULT 13
BZ545807
LOCUS BZ545807 905 bp DNA linear GSS 16-DEC-2002
DEFINITION OGALR72TM ZM2 0.7 1.5 KB Zea mays genomic clone ZMMBMA0089L24,
ACCESSION BZ545807
VERSION BZ545807.1 GI:27099699
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 905)
WhiteLAW, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,
Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.
Consortium for Maize Genomics
Unpublished (2002)
Other GSSs: OGALR72TC
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR
Class: sheared ends.
FEATURES             Location/Qualifiers
     source
     1..905
     /organism="Zea mays"
     /mol_type="genomic DNA"
     /strain="B73"
     /db_xref="taxon:4577"
     /clone="ZMMBMA0089L24"
     /clone_lib="ZM2 0.7 1.5 KB"
     /note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN
Alignment Scores:
Pred. No.:      97      Length:      905
Score:          76.00     Matches:     34
Percent Similarity: 36.28%  Conservative: 7
Best Local Similarity: 30.09% Mismatches: 29
Query Match:      16.20%  Indels:      43
DB:              28      Gaps:         6

US-09-917-376-5 (1-88) x BZ545807 (1-905)

```

methylation filtered genomC DNA library"		
ORIGIN		
Alignment Scores:		
Pred. No.:	112	Length: 1001
Score:	76.00	Matches: 34
Percent Similarity:	36.28%	Conservative: 7
Best Local Similarity:	30.09%	Mismatches: 29
Query Match:	16.20%	Indels: 43
DB:	29	Gaps: 6

US-09-917-376-5 (1-88) x CG270748 (1-1001)

```

QY      2 SerGlyGlyValGlnTyrLysAsnAspSer-----AlaPro 16
Db      418 TCCGAGGCGTCAGCATCCCATACAAGGAGCGCAACAGCCTGCGGAGCTGGCGCC 477
QY      17 Gly-----AspAsnGlnIleLysProGlyLeuGlnLeuValAsn----- 29
Db      478 GCGGCCCGAGCTGAGCAGCTGCGGAGCGGAGCCAGACTGATGCTTATCAACCGTGGCGCG 537
QY      30 Thr-GlySerSerSerValAsp----- 36
Db      538 ACGAGGAAGTTCGTCAACTCCCGGAGATCGTCGGAGCGGTCCAGAAGGCTGGGTTTCGAG 597
QY      37 -----LeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySerSe 53
Db      598 GTATCCCGATCGAGCGCGGTAGAGACCTCAGCGTGGAGGACTTCGCCCGGACCGTCCGAC 657
QY      53 rThrLeuValTyrAsnCysAspTrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPh 73
Db      658 TCGTGCG-----ACGTGCTCATGGGCGCGCACGGGG----- 688
QY      73 eGlySerValAsnProAlaThrProThrAlaAspThr 85
Db      689 -----CCGGCCTCACCAACTTCTTCTCC 712

```

Search completed: May 12, 2004, 11:39:02  
 Job time : 828.547 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 12, 2004, 04:43:36 ; Search time 21.976 Seconds  
(without alignments)  
2222.227 Million cell updates/sec

Title: US-09-917-376-5  
Perfect score: 459  
Sequence: 1 VSGGVKQVKNDSAPGDNQ.....IRASFGSVNPTPATDTYLQ 88

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPRO\_pool/US09917376/runat\_11052004\_114536\_28265/app\_query.fasta\_1.1429  
-DB=Issued Patents NA -QMT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09917376 SCGN 1 1 116 @runat\_11052004\_114536\_28265 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PTCUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	195.5	41.7	2029	US-09-136-574A-46	Sequence 46, Appl
2	194.5	41.0	6416	US-09-136-574A-2	Sequence 2, Appl
3	192.5	41.0	11707	US-09-136-574A-1	Sequence 1, Appl
4	182.5	38.9	2977	US-07-862-588B-1	Sequence 1, Appl
5	165	35.2	1438	US-09-339-159B-3	Sequence 3, Appl
6	165	35.2	1482	US-09-198-956-9	Sequence 9, Appl
7	165	35.2	1482	US-09-198-955A-11	Sequence 11, Appl
8	165	35.2	1482	US-09-694-531-11	Sequence 11, Appl
9	165	35.2	1482	US-09-670-141-9	Sequence 9, Appl
10	165	35.2	1482	US-10-072-152-11	Sequence 11, Appl
11	120.5	25.7	1624	US-07-862-588B-5	Sequence 5, Appl
12	120.5	25.7	1775	US-07-862-588B-5	Sequence 5, Appl

13	114	24.3	486	1	US-08-048-164A-1	Sequence 1, Appl
c 14	114	24.3	486	1	US-08-048-164A-3	Sequence 3, Appl
15	114	24.3	486	1	US-08-460-462-1	Sequence 1, Appl
c 16	114	24.3	486	1	US-08-460-462-3	Sequence 3, Appl
17	114	24.3	486	1	US-08-460-457-1	Sequence 1, Appl
c 18	114	24.3	486	1	US-08-460-457-3	Sequence 3, Appl
19	114	24.3	486	1	US-08-460-458-1	Sequence 1, Appl
c 20	114	24.3	486	1	US-08-460-458-3	Sequence 3, Appl
21	114	24.3	486	2	US-08-460-455-1	Sequence 1, Appl
c 22	114	24.3	486	2	US-08-460-455-3	Sequence 3, Appl
23	114	24.3	486	2	US-08-330-394A-1	Sequence 1, Appl
c 24	114	24.3	486	2	US-08-330-394A-3	Sequence 3, Appl
25	114	24.3	499	3	US-09-006-636-5	Sequence 5, Appl
26	114	24.3	499	3	US-09-006-632-5	Sequence 5, Appl
27	114	24.3	499	4	US-09-325-274-5	Sequence 5, Appl
28	113	24.1	1146	3	US-09-277-716-21	Sequence 21, Appl
29	113	24.1	1146	4	US-09-609-161B-21	Sequence 21, Appl
30	77	16.4	8107	4	US-09-335-586-3	Sequence 3, Appl
31	75.5	16.1	5714	4	US-09-820-312D-393	Sequence 393, App
32	72.5	15.5	20284	4	US-09-526-193A-21	Sequence 21, Appl
33	70	14.9	25165	4	US-09-453-702B-39	Sequence 39, Appl
34	67.5	14.4	2241	2	US-08-838-219B-20	Sequence 20, Appl
35	67.5	14.4	2241	3	US-09-233-336A-20	Sequence 20, Appl
36	67.5	14.4	2241	3	US-09-233-752A-20	Sequence 20, Appl
37	67.5	14.4	2241	3	US-09-402-036-20	Sequence 20, Appl
38	67.5	14.4	2241	4	US-09-904-226-20	Sequence 20, Appl
39	67.5	14.4	2370	2	US-08-838-219B-19	Sequence 19, Appl
40	67.5	14.4	2370	3	US-09-233-336A-19	Sequence 19, Appl
41	67.5	14.4	2370	3	US-09-233-752A-19	Sequence 19, Appl
42	67.5	14.4	2370	3	US-09-402-036-19	Sequence 19, Appl
43	67.5	14.4	2370	4	US-09-904-226-19	Sequence 19, Appl
44	67.5	14.4	118067	4	US-09-497-855A-32	Sequence 32, Appl
45	66	14.1	711	1	US-08-061-092A-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1  
US-09-136-574A-46  
; Sequence 46, Application US/09136574A  
; Patent No. 6294366  
; GENERAL INFORMATION:  
; APPLICANT: Farrington, Graham K.  
; Anderson, Paige  
; Gibbs, Moreland  
; Bergquist, Peter  
; Daniels, Roy  
; Morgan, Hugh W.  
; Williams, Diane P.  
; TITLE OF INVENTION: Compositions and Methods for  
; Treating Cellulose Containing  
; Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Howson and Howson  
STREET: Spring House Corporate Center, P.O. Box 457  
CITY: Spring House  
STATE: PA  
COUNTRY: USA  
ZIP: 19477

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/136.574A  
FILING DATE: 19-Aug-1998  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/932,571  
FILING DATE: September 19, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Bak, Mary E.



```
; STATE: PA
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/136,574A
; FILING DATE: 19-Aug-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/932,571
; FILING DATE: September 19, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: 1997US0001/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11707 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-136-574A-1

Alignment Scores:
Pred. No.: 3,15e-15 Length: 11707
Score: 192.50 Matches: 37
Percent Similarity: 62.35% Conservative: 16
Best Local Similarity: 43.53% Mismatches: 31
Query Match: 41.04% Indels: 1
DB: 3 Gaps: 1

US-09-917-376-5 (1-88) x US-09-136-574A-1 (1-11707)
Qy 4 GlyValIyGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
Db 4038 GGTTTGAAGGTACTATACAAAGCAATGAGACAAAGTGGAGCACAAGTTCTATAAGGCG 4097
Qy 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
Db 4098 TGGTTTAAGATAGTGAATGGAGGAGCAGCAGTGTGTAGTCTTAGCAGGTTTAAGATAAGA 4157
Qy 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63
Db 4158 TACTGGTACACAGTGGATGGTGACAAAGCCACAGAGTGGCGTA---TGTGACTGGGCACAG 4214
Qy 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
Db 4215 ATAGGGCAGCAGTATGTGACATTCAATTTGTGAGCTGAGCAGCGGAGTGAGTGGAGCG 4274
Qy 84 AspThrTyrLeuGln 88
Db 4275 GATTATTACTTGGAG 4289

RESULT 4
US-07-862-588B-1
; Sequence 1, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linea
; APPLICANT: Sch lein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 59167960 No. 5916796disk of No. 5916796th America, Inc.

; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
; REFERENCE/DOCKET NUMBER: 3425.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2977 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Bacillus lautus
; STRAIN: NCIMB 40250
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 677..2776
; OTHER INFORMATION:
; US-07-862-588B-1

Alignment Scores:
Pred. No.: 1,05e-14 Length: 2977
Score: 182.50 Matches: 36
Percent Similarity: 66.27% Conservative: 19
Best Local Similarity: 43.37% Mismatches: 25
Query Match: 38.91% Indels: 3
DB: 2 Gaps: 2

US-09-917-376-5 (1-88) x US-07-862-588B-1 (1-2977)
Qy 6 LysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysProGlyLeu 25
Db 144 CGATGCGATACAGCGCGCGGATACAAATGCGAGCCGACACAGATCAAGCGCTCCTTC 203
Qy 26 GlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArgTyrTrp 45
Db 204 AACATCAAAAACAACCGGTACTTCGGCTGTGATTAAAGCAGCGCTCAAAATCCGCTACTAC 263
Qy 46 PheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAlaMetGly 65
Db 264 TTCACCAAGATGGTTCTCGCGCGGTGAACGGCTGG---ATCGACTGGGCGCAGCTCGGC 320
Qy 66 CysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAlaAspThr 85
Db 321 GGCAGCAACATTCAGATCTCGTTTGGC-----AACCATACTGCACGAATTCGATACG 374
Qy 86 TyrLeuGln 88
Db 375 TACGTGGAG 383
```



```
RESULT 5
US-09-339-159B-3
; Sequence 3, Application US/09339159B
; Patent No. 6566114
; GENERAL INFORMATION:
; APPLICANT: Kauppinen, Markus
; APPLICANT: Schulein, Martin
; APPLICANT: Schnorr, Kirk
; APPLICANT: Andersen, Lene
; APPLICANT: Bjornvad, Mads
; TITLE OF INVENTION: No. 6566114el Mannanases
; FILE REFERENCE: 5440.204-US
; CURRENT APPLICATION NUMBER: US/09/339,159B
; CURRENT FILING DATE: 1999-06-24
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 1438
; TYPE: DNA
; ORGANISM: Bacillus sp. 1633
US-09-339-159B-3

Alignment Scores:
Pred. No.:      8,98e-13      Length:      1438
Score:          165.00      Matches:      35
Percent Similarity: 59.79%      Conservative: 23
Best Local Similarity: 36.08%      Mismatches:  29
Query Match:    35.18%      Indels:      10
DB:              4          Gaps:          2

US-09-917-376-5 (1-88) x US-09-339-159B-3 (1-1438)
QY 1 ValSerGlyVallyValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGln 20
Db 940 GTATCAGGCAATTGGAAGTTGAAATCTACACAGCAATCCTTCAGATACTACTCA 999
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1000 ATCAATCCTCAGTTCAGGTTACTAATACCGAAGCAGTGCATTTGTTGCCAACTC 1059
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1060 ACATTGAGATATTATTATACAGTAGACGACAGAAAGATCAGACCTTCTGG--TGTGAC 1116
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1117 CATGCTGCAATAATCGGCAGTAAACGCGAGCTACACGGAATTTACTTCAAAATGTAAGA 1176
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1177 ACATTGTAAATGAGTTCTCTCAACAATAACGACAGACCTACTCTTGAA 1227

RESULT 6
US-09-198-956-9
; Sequence 9, Application US/09198956
; Patent No. 6165769
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels Erik K.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Schnorr, Kirk
; TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus
; FILE REFERENCE: 5377.200-US
; CURRENT APPLICATION NUMBER: US/09/198,956
; CURRENT FILING DATE: 1998-11-24
; EARLIER APPLICATION NUMBER: 1344/97
; EARLIER FILING DATE: 1997-11-24
; EARLIER APPLICATION NUMBER: 60/067,240
; EARLIER FILING DATE: 1997-12-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSEQ for Windows Version 3.0
US-09-198-956-9

Alignment Scores:
Pred. No.:      9,37e-13      Length:      1482
Score:          165.00      Matches:      35
Percent Similarity: 59.79%      Conservative: 23
Best Local Similarity: 36.08%      Mismatches:  29
Query Match:    35.18%      Indels:      10
DB:              3          Gaps:          2

US-09-917-376-5 (1-88) x US-09-198-956-9 (1-1482)
QY 1 ValSerGlyVallyValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGln 20
Db 991 GTATCAGGCAATTGGAAGTTGAAATCTACACAGCAATCCTTCAGATACTACTCA 1050
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1051 ATCAATCCTCAGTTCAGGTTACTAATACCGAAGCAGTGCATTTGTTGCCAACTC 1110
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1111 ACATTGAGATATTATTATACAGTAGACGACAGAAAGATCAGACCTTCTGG--TGTGAC 1167
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1168 CATGCTGCAATAATCGGCAGTAAACGCGAGCTACACGGAATTTACTTCAAAATGTAAGA 1227
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1228 ACATTGTAAATGAGTTCTCTCAACAATAACGACAGACCTACTCTTGAA 1278

RESULT 7
US-09-198-955A-11
; Sequence 11, Application US/09198955A
; Patent No. 6187580
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongbak, Lars
; TITLE OF INVENTION: No. 6187580el Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/09/198,955A
; CURRENT FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSEQ for Windows Version 4.0
US-09-198-955A-11
; SEQ ID NO 11
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Clostridium thermocellum
US-09-198-955A-11
```

```
Alignment Scores:
Pred. No.: 9.37e-13 Length: 1482
Score: 165.00 Matches: 35
Percent Similarity: 59.79% Conservative: 23
Best Local Similarity: 36.08% Mismatches: 29
Query Match: 35.18% Indels: 10
DB: 3 Gaps: 2

US-09-917-376-5 (1-88) x US-09-198-955A-11 (1-1482)
QY 1 ValSerGlyGlyValIysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 991 GTATCAGGCAATTGAAGGTTGAATCTTACACAGCAATCCTTCAGATACTACTAACTCA 1050
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1051 ATCAATCTCAGTTCAGGTTACTAATACCGAAGCAGTGCATTTGATTTGTCCTCAACTC 1110
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1111 ACATTGAGATATTATTATACAGTAGACGACGACAGAAAGATCAGACCTTCTGG--TGTGAC 1167
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgala 71
Db 1168 CATGCTGCAATAATCGGCAGTAAACGCGAGCTACAACGGAATTACTTCAAAATGTAAAGGA 1227
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1228 ACATTTGTAAAAATGAGTTCTCTCAACAAATAACGACAGACCTTCTGAA 1278

RESULT 8
US-09-694-531-11
; Sequence 11, Application US/09694531
; Patent No. 6368843
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. 6368843el Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/09/694,531
; CURRENT FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 09/198,955
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Clostridium thermocellum
US-09-694-531-11
Alignment Scores:
Pred. No.: 9.37e-13 Length: 1482
Score: 165.00 Matches: 35
Percent Similarity: 59.79% Conservative: 23
Best Local Similarity: 36.08% Mismatches: 29
Query Match: 35.18% Indels: 10
DB: 3 Gaps: 2

US-09-917-376-5 (1-88) x US-09-670-141-9 (1-1482)
QY 1 ValSerGlyGlyValIysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 991 GTATCAGGCAATTGAAGGTTGAATCTTACACAGCAATCCTTCAGATACTACTAACTCA 1050
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1051 ATCAATCTCAGTTCAGGTTACTAATACCGAAGCAGTGCATTTGATTTGTCCTCAACTC 1110
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1111 ACATTGAGATATTATTATACAGTAGACGACGACAGAAAGATCAGACCTTCTGG--TGTGAC 1167
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgala 71
Db 1168 CATGCTGCAATAATCGGCAGTAAACGCGAGCTACAACGGAATTACTTCAAAATGTAAAGGA 1227
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1228 ACATTTGTAAAAATGAGTTCTCTCAACAAATAACGACAGACCTTCTGAA 1278

RESULT 9
US-09-670-141-9
; Sequence 9, Application US/09670141
; Patent No. 6429000
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels Erik K.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Schnorr, Kirk
; TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus
; FILE REFERENCE: 5377.200-US
; CURRENT APPLICATION NUMBER: US/09/670,141
; CURRENT FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: 09/198,956
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-670-141-9
Alignment Scores:
Pred. No.: 9.37e-13 Length: 1482
Score: 165.00 Matches: 35
Percent Similarity: 59.79% Conservative: 23
Best Local Similarity: 36.08% Mismatches: 29
Query Match: 35.18% Indels: 10
DB: 4 Gaps: 2

US-09-917-376-5 (1-88) x US-09-670-141-9 (1-1482)
QY 1 ValSerGlyGlyValIysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 991 GTATCAGGCAATTGAAGGTTGAATCTTACACAGCAATCCTTCAGATACTACTAACTCA 1050
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1051 ATCAATCTCAGTTCAGGTTACTAATACCGAAGCAGTGCATTTGATTTGTCCTCAACTC 1110
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1111 ACATTGAGATATTATTATACAGTAGACGACGACAGAAAGATCAGACCTTCTGG--TGTGAC 1167
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgala 71
Db 1168 CATGCTGCAATAATCGGCAGTAAACGCGAGCTACAACGGAATTACTTCAAAATGTAAAGGA 1227
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1228 ACATTTGTAAAAATGAGTTCTCTCAACAAATAACGACAGACCTTCTGAA 1278
```

1051	ATCAATCCTCAGTTCAAGGTTACTAATACCGGAAGCAGTCGAATTGATTGTCCTCAAACTC	1110
Db		
QY	41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60	
Db	1111 ACATTGAGATAATTATTATACAGTAGACGACAGAGAATCAGACCTTCTGG--TGTGAC 1167	
QY	61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71	
Db	1168 CATGCTGCAATAATCGGCAGTAGCCGAGCTACACGAATTACTTCAAAATGTAAGA 1227	
QY	72 SerPheClySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88	
Db	1228 ACATTTGTAATAATAGATTCTCTCAACAAATAACGACACACCTACTCTTGA 1278	

```

RESULT 10
US-10-072-152-11
; Sequence 11, Application US/10072152
; Patent No. 6677147
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. 6677147el Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/10/072,152
; CURRENT FILING DATE: 2002-02-07
; PRIOR APPLICATION NUMBER: US/09/198,955
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Clostridium thermocellum
US-10-072-152-11

```

[illegible]

```

Db      1111 ACATTGAGATATTATTATACAGTACGACGACAAAGATCAGACCTTCTCG---TGTGAC 1167
QY      61  TtpAlaIaMet-----GlyCysGlyAsnIleArgAla 71
Db      1168 CATGCTGCATATATCGGCAGTACGCGACGACTCAACGGATATTACTCAATGTATAAGGA 1227
QY      72  SerPheGlySerValAsnProIaThrPrOThrAlaAspThrTyrLeuGln 98
Db      1228 ACATTTGTAATAATGAGTTCTCTCAACAAATAACGCGACACCTTACCTTGAA 1278

RESULT 11
US-07-862-588B-6
; Sequence 6, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linnaa
; APPLICANT: Sch lein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5916796o No. 5916796disk of No. 5916796th America, Inc.
; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
; REFERENCE/DOCKET NUMBER: 3425.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1624 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Bacillus lautus
; STRAIN: NCIMB 40250
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 30..1607
; OTHER INFORMATION:
US-07-862-588B-6

Alignment Scores:
Pred. No.: 1,17e-06 Length: 1624
Score: 120.50 Matches: 28
Percent Similarity: 54.12% Conservative: 18
Best Local Similarity: 32.94% Mismatches: 36
Query Match: 25.69% Indels: 3

```

```
DB:                2              Gaps:                2
US-09-917-376-5 (1-88) x US-07-862-588B-6 (1-1624)

Qy      2 SerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     1323 ACGGGAACTTGTTGTCAATACAAAGTTGGCGACACTAGCCGCCAGGATAACCAATG 1382
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy     22 LysProGlyLeuGlnLeuValAenThrGlySerSerValAspLeuSerThrValThr 41
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     1383 AGCCCTTCCTTTTAACAATCAAGAACAACGGGTACAAACCCTGTTAACTGAGCGCCTCAAG 1442
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy     42 ValArgTyTrpPheThrArgAspGlySerSerThrLeuValTyrAsnCysAspTyr 61
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     1443 CTTNNNNNNNNNNNNAAGAAC---GGACCTGCGGATATGAGCTCGATCGACTGG 1499
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy     62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAenProAlaThrPro 81
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     1500 GCCCAATCGCGGAACGAATGTTCTGCTGGCATTC-----GCTAACTTTACCGGGAGT 1553
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy     82 ThrAlaAspThrTyr 86
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     1554 AATACGGATACTTAC 1568
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 12
US-07-862-588B-5
; Sequence 5, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linnae
; APPLICANT: Sch lein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: No. 5916796o No. 5916796disk of No. 5916796th America, Inc.
; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
; REFERENCE/DOCKET NUMBER: 3425.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1775 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Bacillus lautus
```

```
/
/ TELEPHONE: (212) 790-9090
/ TELEFAX: (212) 869-8864/9741
/ TELEX: 66141 PENNIE
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 486 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: unknown
/ MOLECULE TYPE: DNA
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 1..486
/
US-08-048-164A-1

Alignment Scores:
Pred. No.: 1.64e-06 Length: 486
Score: 114.00 Matches: 29
Percent Similarity: 52.58% Conservative: 22
Best Local Similarity: 29.90% Mismatches: 34
Query Match: 24.31% Indels: 12
DB: 1 Gaps: 4

US-09-917-376-5 (1-88) x US-08-048-164A-1 (1-486)
QY 2 SerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21
Db 7 ACATCATCAATGTCAGTTGAATTTTACAACCTCTAACAAATCAGCACAACAACTCAATT 66
QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
Db 67 ACACCAATAATCAAAATTAATAACATCTGACAGTGAATTTAAATTTAAATGACGTAAAA 126
QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61
Db 127 GTTAGATATTATTACACAGTGTGTTACACAGGACAACTTTCTCG---TGTGACCAT 183
QY 62 AlaAlaMetGlyCysGlyAsn-----IleArgAlaSerPhe 73
Db 184 GCTGTGTCATTATTAGGAATAGCTATGTTGATACTAGCAAGAGTGACACAACTTC 243
QY 74 -----GlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 244 GTTAAAGAAACAGCAAGCCCA---ACATCAACCTATGATACATATGTTGAA 291

RESULT 14
US-08-048-164A-3/c
/ Sequence 3, Application US/08048164A
/ Patent No. 5496934
/ GENERAL INFORMATION:
/ APPLICANT: Shoseyov, Oded
/ APPLICANT: Shipiegl, Itai
/ APPLICANT: Goldstein, Marc A.
/ APPLICANT: Doi, Roy H.
/ TITLE OF INVENTION: CELLULOSE BINDING DOMAIN
/ NUMBER OF SEQUENCES: 21
/ CORRESPONDENCE ADDRESS:
/ ADDRESS: PENNIE & EDMONDS
/ STREET: 1155 Avenue of the Americas
/ CITY: New York
/ STATE: New York
/ COUNTRY: U.S.A.
/ ZIP: 10036
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/048,164A
/ FILING DATE: 14-APR-1993
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:

/
/ NAME: Mirock, S. Leslie
/ REGISTRATION NUMBER: 18,872
/ REFERENCE/DOCKET NUMBER: 7809-003
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (212) 790-9090
/ TELEFAX: (212) 869-8864/9741
/ TELEX: 66141 PENNIE
/ INFORMATION FOR SEQ ID NO: 3:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 486 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: unknown
/ MOLECULE TYPE: DNA
/ US-08-048-164A-3

Alignment Scores:
Pred. No.: 1.64e-06 Length: 486
Score: 114.00 Matches: 29
Percent Similarity: 52.58% Conservative: 22
Best Local Similarity: 29.90% Mismatches: 34
Query Match: 24.31% Indels: 12
DB: 1 Gaps: 4

US-09-917-376-5 (1-88) x US-08-048-164A-3 (1-486)
QY 2 SerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21
Db 480 ACATCATCAATGTCAGTTGAATTTTACAACCTCTAACAAATCAGCACAACAACTCAATT 421
QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
Db 420 ACACCAATAATCAAAATTAATAACATCTGACAGTGAATTTAAATTTAAATGACGTAAAA 361
QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61
Db 360 GTTAGATATTATTACACAGTGTGTTACACAGGACAACTTTCTCG---TGTGACCAT 304
QY 62 AlaAlaMetGlyCysGlyAsn-----IleArgAlaSerPhe 73
Db 303 GCTGTGTCATTATTAGGAATAGCTATGTTGATACTAGCAAGAGTGACACAACTTC 244
QY 74 -----GlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 243 GTTAAAGAAACAGCAAGCCCA---ACATCAACCTATGATACATATGTTGAA 196

RESULT 15
US-08-460-462-1
/ Sequence 1, Application US/08460462
/ Patent No. 5670623
/ GENERAL INFORMATION:
/ APPLICANT: Shoseyov, Oded
/ APPLICANT: Shipiegl, Itai
/ APPLICANT: Goldstein, Marc A.
/ APPLICANT: Doi, Roy H.
/ TITLE OF INVENTION: METHODS OF USE OF CELLULOSE BINDING DOMAIN PROTEINS
/ NUMBER OF SEQUENCES: 21
/ CORRESPONDENCE ADDRESS:
/ ADDRESS: PENNIE & EDMONDS
/ STREET: 1155 Avenue of the Americas
/ CITY: New York
/ STATE: New York
/ COUNTRY: U.S.A.
/ ZIP: 10036
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/460,462
/ FILING DATE: concurrently herewith
/ CLASSIFICATION: 435
```

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/048,164  
FILING DATE: 14-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7809-006  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 486 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..486  
US-08-460-462-1

Alignment Scores:  
Pred. No.: 1.64e-06 Length: 486  
Score: 114.00 Matches: 29  
Percent Similarity: 52.58% Conservative: 22  
Best Local Similarity: 29.90% Mismatches: 34  
Query Match: 24.31% Indels: 12  
DB: 1 Gaps: 4

US-09-917-376-5 (1-88) x US-08-460-462-1 (1-486)

Qy	2	SerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle	21
Db	7	ACATCATCAATGTCAGTTGAATTTTACAACTCTACAAATCAGCACAAACAACTCAATT	66
Qy	22	LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr	41
Db	67	ACACCAATATCAAAATTTACTACACATCTGCACGTGATTTAAATTTAAATGACGTAAA	126
Qy	42	ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp	61
Db	127	GTTAGATATTATTACACAAGTGATGGTACACAAGGACAACTTTCTGG--TGTGACCAT	183
Qy	62	AlaAlaMetGlyCysGlyAsn-----IleArgAlaSerPhe	73
Db	184	GCTGTCATATTATTAGGAAATAGCTATGTTGATACACTAGCAAAAGTCACGCAAACTTC	243
Qy	74	-----GlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln	88
Db	244	GTTAAAGAAACAGCAGGCCCA---ACATCAACCTATGATACATATGTTGAA	291

Search completed: May 12, 2004, 15:15:30  
Job time : 26.976 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 12, 2004, 09:26:52 ; Search time 135.119 Seconds  
(without alignments)  
2950.286 Million cell updates/sec

Title: US-09-917-376-5

Perfect score: 459

Sequence: 1 VSGGVKQVKNDSAPGDNQ.....TRASFGSVNATPTTADTYLQ 88

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2941586 seqs, 2264995651 residues

Total number of hits satisfying chosen parameters: 5883172

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_plus\_p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09917376/runat\_11052004\_114537\_28336/app\_query.fasta\_1.1429  
-DB=Published Applications NA -QFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1  
-LOOPL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomsum62  
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09917376 @CGN 1 1 809 @runat\_11052004\_114537\_28336  
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOC=100  
-LONGLOG -DRV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA.\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*
- 14: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*
- 16: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*
- 17: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*
- 18: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*
- 19: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

RESULT 1  
US-09-917-378-2  
; Sequence 2, Application US/09917378  
; Publication No. US20030119093A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: DECKER, STEPHEN R.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT MANNANASE FROM ACIDOTHERMUS CELLULOLYTICUS  
; FILE REFERENCE: 40197.7US01  
; CURRENT APPLICATION NUMBER: US/09/917.378  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 2289  
; TYPE: DNA  
; ORGANISM: Acidothermus cellulolyticus  
US-09-917-378-2

ALIGNMENTS

1	469	100.0	2289	10	US-09-917-378-2	Sequence 2, Appli
2	466	99.4	3687	10	US-09-917-384-2	Sequence 2, Appli
3	466	99.4	3687	10	US-09-917-383-2	Sequence 2, Appli
4	462	98.5	2869	13	US-09-917-376-2	Sequence 2, Appli
5	462	98.5	2869	13	US-10-155-400-2	Sequence 2, Appli
6	257	54.8	2223	15	US-10-156-761-550	Sequence 550, App
7	257	54.8	9025608	15	US-10-156-761-1	Sequence 1, Appli
8	190	40.5	1527	16	US-10-369-493-46838	Sequence 46838, A
9	165	35.2	1438	13	US-10-372-054-3	Sequence 3, Appli
10	165	35.2	1482	13	US-10-655-433-11	Sequence 11, Appl
11	165	35.2	1482	14	US-10-072-152-11	Sequence 11, Appl
12	157	33.5	1314	9	US-09-974-300-684	Sequence 684, App
13	155.5	33.2	4059	10	US-09-784-554B-1	Sequence 1, Appli
14	150.5	32.1	4056	10	US-09-784-554B-3	Sequence 3, Appli
15	130.5	27.8	4452	13	US-10-282-122A-15299	Sequence 15299, A
16	114	24.3	768	15	US-10-261-446-5	Sequence 5, Appli
17	113	24.1	1146	10	US-09-808-898-21	Sequence 21, Appl
18	78.5	16.7	3147	13	US-10-282-122A-13716	Sequence 13716, A
19	76	16.2	96597	12	US-10-052-482-226	Sequence 226, App
20	75.5	16.1	1041	16	US-10-260-238-742	Sequence 742, App
21	75.5	16.1	2091	13	US-10-211-462-222	Sequence 222, App
22	75.5	16.1	3396	17	US-10-398-458-1	Sequence 1, Appli
23	75.5	16.1	4041	9	US-09-747-835A-21	Sequence 21, Appl
24	75.5	16.1	4041	13	US-10-312-312-21	Sequence 21, Appl
25	75.5	16.1	4353	15	US-10-225-567A-405	Sequence 405, App
26	75.5	16.1	4914	9	US-09-974-298-188	Sequence 188, App
27	75.5	16.1	4914	13	US-10-116-802-38	Sequence 38, Appl
28	75.5	16.1	5714	9	US-09-747-835A-19	Sequence 19, Appl
29	75.5	16.1	5714	9	US-09-747-835A-60	Sequence 60, Appl
30	75.5	16.1	5714	13	US-10-312-312-19	Sequence 19, Appl
31	75.5	16.1	5714	13	US-10-312-312-60	Sequence 60, Appl
32	75.5	16.1	5714	15	US-10-037-270-393	Sequence 393, App
33	75.5	16.1	5714	16	US-10-117-722-393	Sequence 393, App
34	75.5	16.1	5773	9	US-09-747-835A-18	Sequence 18, Appl
35	75.5	16.1	5773	13	US-10-312-312-18	Sequence 18, Appl
36	75	16.0	484	13	US-10-424-599-54043	Sequence 54043, A
37	73	15.6	9025608	15	US-10-156-761-1	Sequence 1, Appli
38	72.5	15.5	402	9	US-09-764-860-85	Sequence 85, Appl
39	72.5	15.5	402	15	US-10-074-095-85	Sequence 85, Appl
40	72.5	15.5	9181	16	US-10-212-872-85	Sequence 85, Appl
41	72.5	15.5	20284	13	US-09-984-827-10	Sequence 10, Appl
42	72.5	15.5	20284	13	US-10-617-334-21	Sequence 21, Appl
43	72.5	15.5	20284	16	US-10-452-510-21	Sequence 21, Appl
44	72	15.4	2402	13	US-10-425-114-9462	Sequence 9462, Ap
45	72	15.4	48667	9	US-09-822-268A-3	Sequence 3, Appli

## Alignment Scores:

Pred. No.: 2,21e-56 Length: 2289  
Score: 469.00 Matches: 88  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 10 Gaps: 0

US-09-917-376-5 (1-88) x US-09-917-378-2 (1-2289)

QY 1 ValSerGlyVallyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20  
DB 1363 GGTGCGGTGGGTGAAGTGCAGTACAGAACAAATGATTCGGCGCGCGGTGATAACCG 1422  
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40  
DB 1423 ATCAACCGGGTCTCCAGTTGGTGAATACGGGGTCTGCTCGGTGGATTTGTCGACGGTG 1482  
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
DB 1483 ACGGTGCGGTACTGGTTACCCGGGATGGTGGTCTGCGACACTGGTGTACAACTGTGAC 1542  
QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80  
DB 1543 TGGCGCGGATGGGTGTGGGAATATCCGGCCTCGTTCCGGTCCGGTGAACCCGCGCAG 1602  
QY 81 ProThrAlaAspThrTyrLeuGln 88  
DB 1603 CCGACGGCGGACACCTACCTGCAG 1626

## RESULT 2

US-09-917-384-2  
; Sequence 2, Application US/09917384  
; Publication No. US20030096342A1

## ; GENERAL INFORMATION:

; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: DECKER, STEPHEN R.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT CELLULASE FROM ACIDOTHERMUS  
; FILE REFERENCE: 40170.6US01  
; CURRENT APPLICATION NUMBER: US/09/917,384  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 3687  
; TYPE: DNA  
; ORGANISM: Artificial Sequence

## ; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: Segment of  
; OTHER INFORMATION: Guxa

US-09-917-384-2

## Alignment Scores:

Pred. No.: 1.13e-55 Length: 3687  
Score: 466.00 Matches: 87  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 98.86% Mismatches: 0  
Query Match: 99.36% Indels: 0  
DB: 10 Gaps: 0

US-09-917-376-5 (1-88) x US-09-917-384-2 (1-3687)

QY 1 ValSerGlyVallyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20  
DB 1750 GTGTCGGGTGGGTGAAGTGCAGTACAGAACAAATGATTCGGCGCGCGGTGATAACCG 1809  
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40  
DB 1810 ATCAACCGGGTCTCCAGTTGGTGAATACCGGGTCTGCTCGGTGGATTTGTCGACGGTG 1869

QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
DB 1870 ACGGTGCGGTACTGGTTACCCGGGATGGTGGTCTGCGACACTGGTGTACAACTGTGAC 1929  
QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80  
DB 1930 TGGCGCGGATGGGTGTGGGAATATCCGGCCTCGTTCCGGTCCGGTGAACCCGCGCAG 1989  
QY 81 ProThrAlaAspThrTyrLeuGln 88  
DB 1990 CCGACGGCGGACACCTACCTGCAG 2013

## RESULT 3

US-09-917-383-2  
; Sequence 2, Application US/09917383  
; Publication No. US20030104522A1

## ; GENERAL INFORMATION:

; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: DECKER, STEPHEN R.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT CELLULASE FROM ACIDOTHERMUS  
; FILE REFERENCE: 40170.6US01  
; CURRENT APPLICATION NUMBER: US/09/917,383  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 3687  
; TYPE: DNA  
; ORGANISM: Artificial Sequence

## ; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: Segment of  
; OTHER INFORMATION: Guxa

US-09-917-383-2

## Alignment Scores:

Pred. No.: 1.13e-55 Length: 3687  
Score: 466.00 Matches: 87  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 98.86% Mismatches: 0  
Query Match: 99.36% Indels: 0  
DB: 10 Gaps: 0

US-09-917-376-5 (1-88) x US-09-917-383-2 (1-3687)

QY 1 ValSerGlyVallyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20  
DB 1750 GTGTCGGGTGGGTGAAGTGCAGTACAGAACAAATGATTCGGCGCGCGGTGATAACCG 1809  
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40  
DB 1810 ATCAACCGGGTCTCCAGTTGGTGAATACCGGGTCTGCTCGGTGGATTTGTCGACGGTG 1869  
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
DB 1870 ACGGTGCGGTACTGGTTACCCGGGATGGTGGTCTGCGACACTGGTGTACAACTGTGAC 1929  
QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80  
DB 1930 TGGCGCGGATGGGTGTGGGAATATCCGGCCTCGTTCCGGTCCGGTGAACCCGCGCAG 1989  
QY 81 ProThrAlaAspThrTyrLeuGln 88  
DB 1990 CCGACGGCGGACACCTACCTGCAG 2013

## RESULT 4

US-09-917-376-2  
; Sequence 2, Application US/09917376  
; Publication No. US20040038334A1



```

; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOXYTICUS
; FILE REFERENCE: 40197.4US01
; CURRENT APPLICATION NUMBER: US/09/917,376
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2869
; TYPE: DNA
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (2869)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-09-917-376-2

Alignment Scores:
Pred. No.: 2,98e-55 Length: 2869
Score: 462.00 Matches: 86
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 97.73% Mismatches: 0
Query Match: 98.51% Indels: 0
DB: 13 Gaps: 0

US-09-917-376-5 (1-88) x US-09-917-376-2 (1-2869)
QY 1 ValSerGlyGlyValGlnValGlnThrGlySerSerSerValAspAsnGln 20
Db 2605 GTGTCGGGTGGGGTGGAGGTGAGTATAGGAATAATGATTCCGGCCGGGTGATAATCAG 2664

QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerSerValAspLeuSerThrVal 40
Db 2665 ATCAAGCCGGGTTCGACGGTGGTGAATATCCGGCTCGTTCGGTGAACCCGGCGAGC 2844

QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValThrAsnCysAsp 60
Db 2605 GTGTCGGGTGGGGTGGAGGTGAGTATAGGAATAATGATTCCGGCCGGGTGATAATCAG 2664

QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 2785 TGGCGCGCGATCGGGTGGGGAATATCCGGCTCGTTCGGTGAACCCGGCGAGC 2844

QY 81 ProThrAlaAspThrTyrLeuGln 88
Db 2845 CCGACGGCGGACACCTACCTGCAG 2868

RESULT 5
US-10-155-400-2
; Sequence 2, Application US/10155400
; Publication No. US20030108988A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOXYTICUS
; FILE REFERENCE: NREL 01-36A
; CURRENT APPLICATION NUMBER: US/10/155,400
; CURRENT FILING DATE: 2002-10-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2869
; TYPE: DNA
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:

```

```

; NAME/KEY: modified_base
; LOCATION: (2869)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-10-155-400-2

Alignment Scores:
Pred. No.: 2,98e-55 Length: 2869
Score: 462.00 Matches: 86
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 97.73% Mismatches: 0
Query Match: 98.51% Indels: 0
DB: 15 Gaps: 0

US-09-917-376-5 (1-88) x US-10-155-400-2 (1-2869)
QY 1 ValSerGlyGlyValGlnThrGlySerSerSerValAspAsnGln 20
Db 2605 GTGTCGGGTGGGGTGGAGGTGAGTATAGGAATAATGATTCCGGCCGGGTGATAATCAG 2664

QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerSerValAspLeuSerThrVal 40
Db 2665 ATCAAGCCGGGTTCGACGGTGGTGAATATCCGGCTCGTTCGGTGAACCCGGCGAGC 2724

QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValThrAsnCysAsp 60
Db 2725 ACGGTGCGGTACTGGTTACCCCGGATGGTGGCTCGTCGACACTGGTGACACTGTGAC 2784

QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 2785 TGGCGCGCGATCGGGTGGGGAATATCCGGCTCGTTCGGTGAACCCGGCGAGC 2844

QY 81 ProThrAlaAspThrTyrLeuGln 88
Db 2845 CCGACGGCGGACACCTACCTGCAG 2868

RESULT 6
US-10-156-761-550
; Sequence 550, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 550
; LENGTH: 2223
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2223)
US-10-156-761-550

Alignment Scores:
Pred. No.: 3,69e-26 Length: 2223
Score: 257.00 Matches: 49
Percent Similarity: 71.26% Conservative: 13
Best Local Similarity: 56.32% Mismatches: 25
Query Match: 54.80% Indels: 0
DB: 15 Gaps: 0

```

US-09-917-376-5 (1-88) x US-10-156-761-550 (1-2223)

QY 2 SerGlyGlyValValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21  
Db 1771 TCGGGCGGCTCAAGTCTCTTACAGAACACGACTCTTCGGCCACCACGACGCAATC 1830  
QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41  
Db 1831 CGGCCAGGCTTCGGATCGTCAACCGGACGCGCTCCCTCGACCTGTCCAAAGGTCAAG 1890  
QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerThrLeuValTyrAsnCysAspTrp 61  
Db 1891 GCCCGCTACTACTTCCCGGACACGCGCTCGCCACCGCTGAAGTCTGCGACTAC 1950  
QY 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81  
Db 1951 GCGGCGTGGCTGTTCACACGCTGTAAGTCTGTAACCTGACACGCGCGTGGCG 2010  
QY 82 ThrAlaAspThrTyrLeuGln 88  
Db 2011 GGAGCGGACGCTACCTCGAA 2031

## RESULT 7

US-10-156-761-1

; Sequence 1, Application US/10156761

; Publication No. US20030119018A1

; GENERAL INFORMATION:

; APPLICANT: OMURA, SATOSHI

; APPLICANT: IKEDA, HARUO

; APPLICANT: ISHIKAWA, JUN

; APPLICANT: HORIKAWA, HIROSHI

; APPLICANT: SHIBA, TADAYOSHI

; APPLICANT: SAKAKI, YOSHIYUKI

; APPLICANT: HATTORI, MASAHIRA

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

; FILE REFERENCE: 249-262

; CURRENT APPLICATION NUMBER: US/10/156,761

; CURRENT FILING DATE: 2002-05-29

; PRIOR APPLICATION NUMBER: JP 2001-204089

; PRIOR FILING DATE: 2001-05-30

; PRIOR APPLICATION NUMBER: JP 2001-272697

; PRIOR FILING DATE: 2001-08-02

; NUMBER OF SEQ ID NOS: 15109

; SEQ ID NO 1

; LENGTH: 9025608

; TYPE: DNA

; ORGANISM: Streptomyces avermitilis

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (4187715)

; OTHER INFORMATION: a, t, c, g, other or unknown

US-10-156-761-1

Alignment Scores:  
Pred. No.: 2,72e-21 Length: 9025608  
Score: 257.00 Matches: 49  
Percent Similarity: 71.26% Conservative: 13  
Best Local Similarity: 56.32% Mismatches: 25  
Query Match: 54.80% Indels: 0  
DB: 15 Gaps: 0

US-09-917-376-5 (1-88) x US-10-156-761-1 (1-9025608)

QY 2 SerGlyGlyValValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21  
Db 706216 TCGGGCGGCTCAAGTCTCTTACAGAACACGACTCTTCGGCCACCACGACGCAATC 706275  
QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41  
Db 706276 CGGCCAGGCTTCGGATCGTCAACCGGACGCGCTCCCTCGACCTGTCCAAAGGTCAAG 706335  
QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerThrLeuValTyrAsnCysAspTrp 61  
Db 706336 GCGCGCTACTACTTCCCGGACACGCGCTCGCCACCGCTGAAGTCTGCGACTAC 706455

Db 706336 GCGCGCTACTACTTCCCGGACACGCGCTCGCCACCGCTGAAGTCTGCGACTAC 706395  
QY 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81  
Db 706396 GCGGCGTGGCTGTTCACACGCTGAGTCTGACCTACCTCGACCGCCGTCGCG 706455  
QY 82 ThrAlaAspThrTyrLeuGln 88  
Db 706456 GGAGCGGACGCTACCTCGAA 706476

## RESULT 8

US-10-369-493-46838

; Sequence 46838, Application US/10369493

; Publication No. US20030233675A1

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493

; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039

; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 46838

; LENGTH: 1527

; TYPE: DNA

; ORGANISM: Bacillus subtilis

US-10-369-493-46838

Alignment Scores:  
Pred. No.: 8.02e-17 Length: 1527  
Score: 190.00 Matches: 35  
Percent Similarity: 65.12% Conservative: 21  
Best Local Similarity: 40.70% Mismatches: 26  
Query Match: 40.51% Indels: 4  
DB: 16 Gaps: 2

US-09-917-376-5 (1-88) x US-10-369-493-46838 (1-1527)

QY 4 GlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23  
Db 1084 GGTATTTCTGTACAGTACACAGCAGCGGATGGAGTATGAACACGACCAATCGTCGG 1143  
QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43  
Db 1144 CAGCTTCAATAAATAAATAACGCGCAATACACGCTTGTATTTAAAGATGTCTGCTGCCGT 1203  
QY 44 TyrTrpPhe---ThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAla 62  
Db 1204 TACTGGTATAAGCAAAACAAAGGCAAAAC-----TTTGACTGTGACTACGCG 1254  
QY 63 AlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThr 82  
Db 1255 CAGATTGGATGCGGCAATGTGACACACAAGTTTGTGCTGCAATAAACCAAGCAAGGT 1314

QY 83 AlaAspThrTyrLeuGln 88

Db 1315 GCAGATACCTATCTCGAA 1332

## RESULT 9

US-10-372-054-3

; Sequence 3, Application US/10372054

; Publication No. US20030203466A1

; GENERAL INFORMATION:

; APPLICANT: Kauppinen, Markus

; APPLICANT: Schuelein, Martin

; APPLICANT: Schnorr, Kirk

; APPLICANT: Andersen, Lene

```

; APPLICANT: Bjornvad, Mads
; TITLE OF INVENTION: Novel Mannanases
; FILE REFERENCE: 5440.204-US
; CURRENT APPLICATION NUMBER: US/10/372,054
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: US/09/339,159B
; PRIOR FILING DATE: 1999-06-24
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 3
; LENGTH: 1438
; TYPE: DNA
; ORGANISM: Bacillus sp. 1633
US-10-372-054-3

Alignment Scores:
Pred. No.: 2,726-13 Length: 1438
Score: 165.00 Matches: 35
Percent Similarity: 59.79% Conservative: 23
Best Local Similarity: 36.08% Mismatches: 29
Query Match: 35.18% Indels: 10
DB: 13 Gaps: 2

US-09-917-376-5 (1-88) x US-10-372-054-3 (1-1438)
QY 1 ValSerGlyGlyValValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 940 GTATCAGCAATTGCAAGTTGAATCTTACACAGCAATCCTTCAGATACTACTAACTCA 999
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1000 ATCAATCCTCAGTTCAAGTTACTAATACCGGAAGCAGTGCAATTGATTTGCCAAACTC 1059
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1060 ACATTGAGATATTATTATACAGTAGACGACGAGAAAGATCAGACCTTCTGG---TGTGAC 1116
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1117 CATGCTGCAATTAATCGGCAGTAACGGCAGTACACAGGAATTAATTCATAATGTAAAGGA 1176
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1177 ACATTGTAAATAGTTCCTCAACAAATAACGACGACACCTACTCTTGAA 1227

RESULT 10
US-10-655-433-11
; Sequence 11, Application US/10655433
; Publication No. US20040067572A1
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. US20040067572A1el Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/10/655,433
; CURRENT FILING DATE: 2003-09-04
; PRIOR APPLICATION NUMBER: US/09/198,955A
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217

; APPLICANT: Bjornvad, Mads
; TITLE OF INVENTION: Novel Mannanases
; FILE REFERENCE: 5440.204-US
; CURRENT APPLICATION NUMBER: US/10/372,054
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: US/09/339,159B
; PRIOR FILING DATE: 1999-06-24
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 3
; LENGTH: 1438
; TYPE: DNA
; ORGANISM: Clostridium thermocellum
US-10-655-433-11

Alignment Scores:
Pred. No.: 2,846-13 Length: 1482
Score: 165.00 Matches: 35
Percent Similarity: 59.79% Conservative: 23
Best Local Similarity: 36.08% Mismatches: 29
Query Match: 35.18% Indels: 10
DB: 13 Gaps: 2

US-09-917-376-5 (1-88) x US-10-655-433-11 (1-1482)
QY 1 ValSerGlyGlyValValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 991 GTATCAGCAATTGCAAGTTGAATCTTACACAGCAATCCTTCAGATACTACTAACTCA 1050
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1051 ATCAATCCTCAGTTCAAGTTACTAATACCGGAAGCAGTGCAATTGATTTGCCAAACTC 1110
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1111 ACATTGAGATATTATTATACAGTAGACGACGAGAAAGATCAGACCTTCTGG---TGTGAC 1167
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1168 CATGCTGCAATTAATCGGCAGTAACGGCAGTACACAGGAATTAATTCATAATGTAAAGGA 1227
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1228 ACATTGTAAATAGTTCCTCAACAAATAACGACGACACCTACTCTTGAA 1278

RESULT 11
US-10-072-152-11
; Sequence 11, Application US/10072152
; Publication No. US20020142438A1
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. US20020142438A1el Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/10/072,152
; CURRENT FILING DATE: 2002-02-07
; PRIOR APPLICATION NUMBER: US/09/198,955
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217

```

```

; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Clostridium thermocellum
US-10-072-152-11

Alignment Scores:
Pred. No.: 2,846-13 Length: 1482
Score: 165.00 Matches: 35
Percent Similarity: 59.79% Conservative: 23
Best Local Similarity: 36.08% Mismatches: 29
Query Match: 35.18% Indels: 10
DB: 14 Gaps: 2

US-09-917-376-5 (1-88) x US-10-072-152-11 (1-1482)

QY 1 ValSerGlyValLeuValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 991 GTATCAGCAATTGAAGTTGAATTTCTACACAGCAATCTTCAGATACTACTACTCA 1050
QY 21 IleLysProGlyLeuLlnValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1051 ATCAATCTCAGTTCAAGTTTACTAATACCGAAGCAGTGAATTTGTCCTCAAACTC 1110
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1111 ACATTGAGATATTATATACAGTAGCAGGACAGAAAGATCAGACCTCTCGG---TGTGAC 1167
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1168 CATGCTGCAATAATCGGCAGTAACGCACTACAACGGAATTACTTCAAAATGTAAAGGA 1227
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1228 ACATTGTAAATAATGAGTTCTCTCAACAAATAACGACAGACACCTACCTTGAA 1278

RESULT 12
US-09-974-300-684
; Sequence 684, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: 10085-500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 684
; LENGTH: 1314
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-684

Alignment Scores:
Pred. No.: 3,346-12 Length: 1314
Score: 157.00 Matches: 32
Percent Similarity: 56.47% Conservative: 16
Best Local Similarity: 37.65% Mismatches: 33
Query Match: 33.48% Indels: 4
DB: 9 Gaps: 2

US-09-917-376-5 (1-88) x US-09-974-300-684 (1-1314)

; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Clostridium thermocellum
US-10-072-152-11

Alignment Scores:
Pred. No.: 2,846-13 Length: 1482
Score: 165.00 Matches: 35
Percent Similarity: 59.79% Conservative: 23
Best Local Similarity: 36.08% Mismatches: 29
Query Match: 35.18% Indels: 10
DB: 14 Gaps: 2

US-09-917-376-5 (1-88) x US-10-072-152-11 (1-1482)

QY 1 ValSerGlyValLeuValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 991 GTATCAGCAATTGAAGTTGAATTTCTACACAGCAATCTTCAGATACTACTACTCA 1050
QY 21 IleLysProGlyLeuLlnValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1051 ATCAATCTCAGTTCAAGTTTACTAATACCGAAGCAGTGAATTTGTCCTCAAACTC 1110
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1111 ACATTGAGATATTATATACAGTAGCAGGACAGAAAGATCAGACCTCTCGG---TGTGAC 1167
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1168 CATGCTGCAATAATCGGCAGTAACGCACTACAACGGAATTACTTCAAAATGTAAAGGA 1227
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1228 ACATTGTAAATAATGAGTTCTCTCAACAAATAACGACAGACACCTACCTTGAA 1278

RESULT 13
US-09-784-554B-1
; Sequence 1, Application US/09784554B
; Publication No. US20030032162A1
; GENERAL INFORMATION:
; APPLICANT: Schnorr, Kirk
; APPLICANT: Jorgensen, Per Lina
; APPLICANT: Schuelein, Martin
; TITLE OF INVENTION: FAMILY 44 XYOLOGUCANASES
; FILE REFERENCE: 10017-200-US
; CURRENT APPLICATION NUMBER: US/09/784,554B
; CURRENT FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 1
; LENGTH: 4059
; TYPE: DNA
; ORGANISM: Paenibacillus polymyxa
US-09-784-554B-1

Alignment Scores:
Pred. No.: 2,5e-11 Length: 4059
Score: 155.50 Matches: 27
Percent Similarity: 56.47% Conservative: 21
Best Local Similarity: 31.76% Mismatches: 36
Query Match: 33.16% Indels: 1
DB: 10 Gaps: 1

US-09-917-376-5 (1-88) x US-09-784-554B-1 (1-4059)

QY 4 GlyValysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
Db 3598 GGACTGCTGCTCCAATATCGCACAGCAGATACTAAGGTGAACGATAATCACTCAATCCG 3657
QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
Db 3658 CAATTCCAATTTGTAAACAAGGACACACCTCCATCCGATCCAGATGGAATTCGC 3717
QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63
Db 3718 TACTACTACAAATCGACGCGTGACCGTGAGCAG--ACTTTCAACTGCGACTATCGGACG 3774
QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
Db 3775 CTGAGCTGCTCAAAAGCTGAACGGTAAACTGTTAAATGGAGAGGCTGCAACCGGTGCC 3834
QY 84 AspThrTyrLeuGln 88
Db 3835 GATTATTATTGGAA 3849

RESULT 14
US-09-784-554B-3
; Sequence 3, Application US/09784554B

```

```
; Publication No. US20030032162A1
; GENERAL INFORMATION:
; APPLICANT: Schorr, Kirk
; APPLICANT: Jorgensen, Per Lina
; APPLICANT: Schulein, Martin
; TITLE OF INVENTION: FAMILY 44 XYLOGLYCANASES
; FILE REFERENCE: 1001760-US
; CURRENT APPLICATION NUMBER: US/09/784,554B
; CURRENT FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 4056
; TYPE: DNA
; ORGANISM: Paenibacillus polymyxa
US-09-784-554B-3

Alignment Scores:
Pred. No.: 1.29e-10 Length: 4056
Score: 150.50 Matches: 27
Percent Similarity: 56.47% Conservative: 21
Best Local Similarity: 31.76% Mismatches: 36
Query Match: 32.09% Indels: 1
DB: 10 Gaps: 1

US-09-917-376-5 (1-88) x US-09-784-554B-3 (1-4056)
Qy 4 GlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
Db 3595 GGATGGTGGTCCATGATCGCACGGGATACAAATGATGACGACATCATTGAACCCG 3654
Qy 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
Db 3655 CATTTCCAAATTTTAAATAAGGTACAAATCTCCGTACCGATCAACGAGTTGAAATTCG 3714
Qy 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63
Db 3715 TACTACTACGATCGACGGTGCACCGTGCAGCAG---ACATTCACACTGCGACTATCGCGTG 3771
Qy 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
Db 3772 CTGAGCTCTCGAAGTGAATGGTAAAGTGGTAAATGATGATAAAGCTGCAACCGGTGCT 3831
Qy 84 AspThrTyrLeuGln 88
Db 3832 GATTATTATTGGAA 3846

RESULT 15
US-10-282-122A-15299
; Sequence 15299, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727

; PRIORITY FILING DATE: 2000-05-26
; PRIORITY APPLICATION NUMBER: 60/230,335
; PRIORITY FILING DATE: 2000-09-06
; PRIORITY APPLICATION NUMBER: 60/230,347
; PRIORITY FILING DATE: 2000-09-09
; PRIORITY APPLICATION NUMBER: 60/242,578
; PRIORITY FILING DATE: 2000-10-23
; PRIORITY APPLICATION NUMBER: 60/253,625
; PRIORITY FILING DATE: 2000-11-27
; PRIORITY APPLICATION NUMBER: 60/257,931
; PRIORITY FILING DATE: 2000-12-22
; PRIORITY APPLICATION NUMBER: 60/267,636
; PRIORITY FILING DATE: 2001-02-09
; PRIORITY APPLICATION NUMBER: 60/269,308
; PRIORITY FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 15299
; LENGTH: 4452
; TYPE: DNA
; ORGANISM: Clostridium acetobutylicum
US-10-282-122A-15299

Alignment Scores:
Pred. No.: 1.04e-07 Length: 4452
Score: 130.50 Matches: 30
Percent Similarity: 55.79% Conservative: 23
Best Local Similarity: 31.58% Mismatches: 33
Query Match: 27.83% Indels: 9
DB: 13 Gaps: 2

US-09-917-376-5 (1-88) x US-10-282-122A-15299 (1-4452)
Qy 2 SerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21
Db 100 AATTCGGGTGTTCAATATACAAATTTGCTGTATACAAATATAGTACCAACCATGAATCTATT 159
Qy 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
Db 160 GCTCCTAAATTTAAATCAAAATATAATCTGGAGCACCTTTAGATTATTAACAACCTTAAAA 219
Qy 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61
Db 220 TTAAGATACATATTTTACAGCTGATGCTAGTACTCAGGATGAAAATTTTGG---TGCACCAT 276
Qy 62 AlaAlaMetGlyCysGly-----AsnIleArgAlaSerPhe 73
Db 277 GCTGTATGCTTAATGTTTATACTACCAACCAATTAACAGTAATGTAGTGGTACTTTT 336
Qy 74 GlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 337 GTAGCTATGGATAATGCAACAGCTACTGCTGCTGATCATTTATTTGAG 381

Search completed: May 12, 2004, 18:28:26
Job time : 827.119 secs
```

DATE

5/14/07

APPLICATION NUMBER

09/917,376

DOC CODE

\_\_\_\_\_

DOC DATE

10/6/05

DELIVER THE ATTACHED FILE/DOCUMENT TO THE TC  
SCANNING CENTER

CONTRACTOR: THE ATTACHED FILE/DOCUMENT MUST BE  
INDEXED AND SCANNED INTO IFW WITHIN 8 WORK HOURS;  
UPLOADING OF THE SCANNED IMAGES SHOULD OCCUR NO  
LATER THAN 16 WORK HOURS  
FOLLOWING RECEIPT OF THIS REQUEST

AFTER SCANNING, ORIGINAL DOCUMENTS SHOULD BE BOXED IN  
ACCORDANCE WITH INSTRUCTIONS